

**Transformações em dados compostacionais  
para a aplicação da análise de  
componentes principais**

Ricardo Matioli Messias

DISSERTAÇÃO APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO TÍTULO  
DE  
MESTRE EM ESTATÍSTICA

Programa: Mestrado em Estatística

Orientadora: Prof. Dr<sup>a</sup>. Lúcia Pereira Barroso

São Paulo, Abril de 2016

# **Transformações em dados composicionais para a aplicação da análise de componentes principais**

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 29/04/2016. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof<sup>a</sup>. Prof. Dr<sup>a</sup>. Lúcia Pereira Barroso - IME-USP
- Prof<sup>a</sup>. Dr<sup>a</sup>. Denise Aparecida Botter - IME-USP
- Prof. Dr. Casimiro Jayme Alfredo Sepulveda Munita - IPEN-USP

# Agradecimentos

Muitas pessoas passam por nossas vidas e acabam deixando um pouco das suas próprias vidas dentro das nossas. Aqui agradeço a algumas dessas pessoas que passaram pela minha vida e me ajudaram a chegar até aqui.

Primeiramente gostaria de agradecer meu pai **Ademir Messias** e minha mãe **Ana Cecilia Matioli Messias**, professores, por sempre me incentivarem a estudar e me darem todo o suporte necessário.

Ainda primeiramente nessa minha nova fase de vida agradeço minha esposa, companheira, amiga, estatística e cozinheira **Jiang Pu** por sempre me incentivar e me ajudar a superar os momentos mais difíceis da minha vida e nunca me deixar desistir deste trabalho sempre me lembrando com a frase "E sua dissertação?".

Agradeço a minha orientadora **Lúcia Pereira Barroso** por todas as conversas, incentivos, auxílios e lições passadas desde os tempos de graduação até hoje.

Agradeço também aos meus chefes na *Boa Vista SCPC* **Silvo Watanabe**, **Helena Utsumi Okai** e **Ana Gabriela de Oliveira** por todo o apoio e compreensão nos meus horários, permitindo e incentivando que eu assistisse as aulas do mestrado; sem isso teria sido impossível eu chegar até aqui. Também ao meu atual chefe **Carlos Lira** no *Banco Itaú-Unibanco* por compreender as minhas ausências para finalizar o processo de escrita desta dissertação.

Agradeço a todos do time de Basebol e Softbol do IME-USP pelos momentos de descontração e diversão que tivemos ao longo de todos esses anos.



# Resumo

Messias, R. M **Transformações em dados compositionais para a aplicação da análise de componentes principais.** 2016. 114 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

A análise de dados compositionais está sendo amplamente utilizada nas diversas áreas do conhecimento como por exemplo na análise de sedimentos rochosos, na comparação de diferentes células e até na análise criminalística na comparação de evidências de crimes. Durante a história da análise deste tipo de dados existiram muitos tipos de ajustes utilizados para contornar o problema da soma constante das variáveis e ainda hoje não temos um consenso de qual a melhor solução a ser utilizada. Neste trabalho, temos como objetivo a enumeração das 7 transformações que mais foram utilizadas ao longo do tempo e suas vantagens e desvantagens. A análise de componentes principais foi escolhida para o comparativo destas transformações. Fizemos a aplicação destas transformações em três bancos de dados reais com características diferentes entre si, comparamos os resultados e analisamos qual das transformações apresentou o melhor desempenho em cada base de dados. Os critérios de comparação foram o percentual da variância explicada, as variáveis que foram mais importantes para a primeira componente principal, cargas das variáveis nas componentes principais mais importantes assim como suas correlações com as variáveis. Também, simulamos quatro estruturas de bases de dados compositionais para avaliar o desempenho das transformações. Para essas comparações e simulações, foram desenvolvidas algumas funções, utilizando o *software* estatístico R, que visam facilitar a comparação entre as sete transformações, assim auxiliando na escolha de qual das transformações melhor se adapta aos dados. Pelos resultados obtidos notamos que: nas bases de dados reais, os resultados das explicações da variância das transformações são similares e as transformações Ref e Alr mostram melhores desempenhos que as demais; nas quatro estruturas simuladas as transformações Ref e Alr também possuem os melhores resultados na explicação da variância e a interpretação de suas componentes principais são parecidas, assim como as transformações Trad, Log e Clr. Com isso notamos que independentemente da aplicação do logaritmo nas transformações Alr e Log elas apresentaram resultados muitos similares às transformações Ref e Trad, respectivamente, tanto na explicação da variância como na interpretação das componentes principais.

**Palavras-chave:** *Additive Log-Ratio, Center Log-Ratio, Componentes Principais, Dados Composicionais, Transformação Esférica, Transformação Log.*

# Abstract

Messias, R. M **Transformations in compositional data for application of principal components analysis.** 2016. 114 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

The compositional data analysis is being widely used in several areas of knowledge such as the analysis of rocky sediments, to compare different biological cells and even in forensic analysis to compare crimes evidences. During the history of the analysis of such data, to circumvent the problem of variable's constant sum were used many types of adjustments. Until now, we do not have a consensus in which is the best solution to be used in this cases. In this paper, we aim to enunciate seven transformations that most were used over time and their advantages and disadvantages. The principal component analysis was chosen for the comparison of these transformations. We applied this transformations in three real databases with different characteristics, we hope to compare the results and analyze which transformation have the best performance in each database. The comparison criteria were the percentage of explained variance, the variables that were most important to the first principal component, variable's loads in the most important principal components as well their correlation with the variables. We also simulated four compositional data bases structures to evaluate the performance of the transformations. For these comparisons and simulations were developed some functions, using the statistical software R, to facilitate comparison between the seven transformations, thus assisting in choosing which of the best transformation fits to the data. From the results we note that: for the real databases, the results of the variance explanation of all transformations are similar, thus Ref and Alr transformations show better performances than the others; in the four simulated structures the Ref and Alr transformations also have the best results in the variance explanation and interpretation of its main components are similar, as well as the transformations Trad, Log and Clr. Thus we note that independently of applying logarithm in and Log and Alr transformations they present very similar results as Ref and Trad transformations, respectively, both in variance explanation and in the interpretation of the principal components.

**Keywords:** *Additive Log-Ratio, Center Log-Ratio, Compositional Data, Log Transformation, Principal Component Analysis, Spherical Transformation.*



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Estrutura de Dados . . . . .	1
1.2	Contexto Histórico . . . . .	2
1.3	Estrutura da Dissertação . . . . .	2
<b>2</b>	<b>Transformações</b>	<b>3</b>
2.1	Análise Tradicional (Trad) . . . . .	3
2.2	Exclusão de uma Variável (Exc) . . . . .	4
2.3	Variável de Referência (Ref) . . . . .	4
2.4	Transformação Log (Log) . . . . .	5
2.5	<i>Additive Log-Ratio</i> (Alr) . . . . .	5
2.6	<i>Center Log-Ratio</i> (Clr) . . . . .	6
2.7	Transformação Esférica (Esf) . . . . .	6
<b>3</b>	<b>Base de Dados e Análises</b>	<b>9</b>
3.1	Critérios . . . . .	9
3.2	Base 1 . . . . .	10
3.3	Base 2 . . . . .	13
3.4	Base 3 . . . . .	16
<b>4</b>	<b>Simulação</b>	<b>21</b>
4.1	Dirichlet . . . . .	21
4.2	Uniforme . . . . .	22
4.3	Normal . . . . .	22

4.4	Critérios . . . . .	23
4.5	Simulação 1 . . . . .	24
4.5.1	Dirichlet . . . . .	24
4.5.2	Uniforme . . . . .	25
4.5.3	Normal . . . . .	26
4.6	Simulação 2 . . . . .	27
4.6.1	Dirichlet . . . . .	27
4.6.2	Uniforme . . . . .	28
4.6.3	Normal . . . . .	29
4.7	Simulação 3 . . . . .	30
4.7.1	Dirichlet . . . . .	30
4.7.2	Uniforme . . . . .	31
4.7.3	Normal . . . . .	32
4.8	Simulação 4 . . . . .	34
4.8.1	Dirichlet . . . . .	34
4.8.2	Uniforme . . . . .	35
4.8.3	Normal . . . . .	36
4.9	Resumo das Simulações . . . . .	37
<b>5</b>	<b>Conclusões</b>	<b>39</b>
<b>A</b>	<b>Resultados da Análise de Componentes Principais</b>	<b>41</b>
A.1	Análise base 1 . . . . .	42
A.1.1	Análise Tradicional (Trad) . . . . .	42
A.1.2	Exclusão de uma Variável (Exc) . . . . .	44
A.1.3	Variável de Referência (Ref) . . . . .	46
A.1.4	Transformação Log (Log) . . . . .	48
A.1.5	<i>Additive Log-Ratio</i> (Alr) . . . . .	50
A.1.6	<i>Center Log-Ratio</i> (Clr) . . . . .	52
A.1.7	Transformação Esférica (Esf) . . . . .	54
A.2	Análise base 2 . . . . .	56

A.2.1	Análise Tradicional (Trad) . . . . .	56
A.2.2	Exclusão de uma Variável (Exc) . . . . .	58
A.2.3	Variável de Referência (Ref) . . . . .	60
A.2.4	Transformação Log (Log) . . . . .	62
A.2.5	<i>Additive Log-Ratio</i> (Alr) . . . . .	64
A.2.6	<i>Center Log-Ratio</i> (Clr) . . . . .	66
A.2.7	Transformação Esférica (Esf) . . . . .	68
A.3	Análise base 3 . . . . .	70
A.3.1	Análise Tradicional (Trad) . . . . .	70
A.3.2	Exclusão de uma Variável (Exc) . . . . .	73
A.3.3	Variável de Referência (Ref) . . . . .	76
A.3.4	Transformação Log (Log) . . . . .	79
A.3.5	<i>Additive Log-Ratio</i> (Alr) . . . . .	82
A.3.6	<i>Center Log-Ratio</i> (Clr) . . . . .	85
A.3.7	Transformação Esférica (Esf) . . . . .	88
<b>B</b>	<b>Funções R</b>	<b>91</b>
<b>C</b>	<b>Bases de Dados Utilizadas</b>	<b>111</b>
C.1	Base 1 . . . . .	112
C.2	Base 2 . . . . .	115
C.3	Base 3 . . . . .	116
	<b>Referências Bibliográficas</b>	<b>119</b>



# Capítulo 1

## Introdução

### 1.1 Estrutura de Dados

Dados compostacionais possuem uma estrutura muito particular na qual as variáveis têm soma constante, como por exemplo quando se quer analisar rochas encontradas em diferentes regiões, utilizando a sua composição mineral. A participação de cada mineral na massa da rocha é considerada como uma variável (componente) e a soma destas é igual a massa total da rocha. Além disso, quando se comparam rochas de massas diferentes temos que utilizar uma mesma escala para as variáveis, sendo o usual utilizar o percentual daquele mineral na composição total ao invés da massa daquele mineral presente na rocha. Sendo assim, a restrição de soma constante é forte, pois temos que a soma do percentual das variáveis de todas as rochas será igual a 1.

Por exemplo, imaginem que um geólogo encontra rochas em diversos lugares do mundo que parecem similares. Poderemos compará-las utilizando seus componentes manganês, silício, ferro e outros minerais. Para isso é medido o quanto de cada um desses minerais está presente nas rochas encontradas. Suponha-se que temos 2 rochas com as seguintes composições: (22%, 40%, 15%, 23%) e (17%, 35%, 19%, 29%) de manganês, silício, ferro e outros minerais, respectivamente. Para analisar esse tipo de dados temos alguns problemas estruturais, pois pela soma constante a correlação entre as variáveis é alta uma vez que a soma de todas é restrita a 1 e não é a real relação entre as variáveis. Essa estrutura pode ocasionar erros de interpretação e devido a isso as variáveis não possuem distribuições normais.

Seja  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_p)$  uma matriz  $n \times p$  e  $\mathbf{x}_i$  um vetor  $n \times 1$ , com  $i = 1, 2, \dots, p$ ,  $n$  sendo o número de observações da amostra e  $p$  sendo o número de variáveis mensuradas. Essa estrutura de dados está definida em um espaço amostral diferente do usual espaço amostral euclidiano. Ela está definida no espaço simplex de dimensão  $p - 1$  que contém todas as  $p$  partes da composição dado que a soma das variáveis é constante.

Pode-se encontrar essa estrutura de dados em praticamente todos os ramos da ciência,

como por exemplo, na biologia no estudo e comparação de células, estudo da composição da poluição, em geologia no estudo de rochas e solos, em arqueologia no estudo de comparação de matérias encontradas em diferentes sítios arqueológicos, na criminalística no estudo dos vestígios deixados pelo criminoso, entre outros.

## 1.2 Contexto Histórico

A ideia de que essa estrutura de dados teria que ser analisada de uma forma diferente começou com Pearson (1896), quando ele alertou a comunidade de que as somas constantes trariam problemas na correlação entre as variáveis pela falta de linearidade da relação entre elas. Somente quase 50 anos depois, estudos na área de geologia por Chayes (1956, 1960) e em biologia por Mosimann (1962) mostraram que a utilização de técnicas usuais a essa estrutura de dados levaria a problemas de interpretações dos resultados da análise de correlação entre as variáveis.

Já Butler (1979), Chayes (1971), Chayes e Kruskal (1966), Chayes e Trochimczyk (1978), Darroch e James (1974) utilizaram a análise multivariada normal usual e propuseram a utilização de correções posteriores para a interpretação dos resultados.

Quase 100 anos depois de Pearson alertar sobre o problema da soma constante começaram a nascer as técnicas de análise em dados compostacionais do modo como conhecemos hoje com Aitchison e Shen (1980) e Aitchison (1982, 1983, 1984, 1985, 1986).

## 1.3 Estrutura da Dissertação

No Capítulo 2 apresentamos as definições básicas das 7 transformações usualmente utilizadas em dados compostacionais. São elas: análise tradicional (Trad), exclusão de uma variável (Exc), variável de referência (Ref), transformação log (Log), *additive log-ratio* (Alr), *center log-ratio* (Clr) e finalmente transformação esférica (Esf).

No Capítulo 3 comparamos as 7 transformações utilizando a análise de componentes principais aplicadas em 3 bases de dados. No 4 apresentamos os resultados das simulações. Finalmente, no Capítulo 5 enunciamos as conclusões obtidas nesta dissertação.

No Apêndice A apresentamos todos os resultados obtidos na análise de componentes principais para cada transformação em cada base de dados analisada. Já no Apêndice B, apresentamos as funções desenvolvidas no *software* estatístico R, para a aplicação das 7 transformações em análise de componentes principais e a visualização de seus resultados. Por último no Apêndice C estão as 3 bases utilizadas para as análises do Capítulo 3.

# Capítulo 2

## Transformações

Existem várias abordagens para tentarmos evitar o efeito da soma das variáveis ser restrita. Neste trabalho enunciamos e discutimos algumas delas, que são facilmente interpretáveis. As transformações têm funções inversas facilmente calculadas, com exceção da transformação esférica que possui uma aplicação específica para bases com muitas observações com zeros, e por isso a necessidade de mantê-la no estudo.

### 2.1 Análise Tradicional (Trad)

A análise tradicional consiste em se considerar que os dados são provenientes de variáveis independentes e com distribuição normal, como comumente são consideradas as variáveis em análises multivariadas.

Ela foi utilizada principalmente pelo desconhecimento do efeito que a soma constante tem, naquelas técnicas que exigem a suposição de normalidade dos dados, como regressões lineares, análise discriminante entre outras.

É usual padronizar as variáveis antes de se aplicar as técnicas multivariadas, ou seja, considerar a matriz de dados  $\mathbf{Z}$  dada por:

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_p),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{z}_i$  é  $z_{ki} = \frac{x_{ki} - \bar{x}_i}{s_i}$ ,  $\bar{x}_i$  é a média amostral e  $s_i$  o desvio padrão amostral da variável  $i$ , com  $i = 1, 2, \dots, p$ .

Esse tipo de abordagem pode indicar relações entre as variáveis derivadas do fato de a soma ser constante e não porque as variáveis têm realmente relação entre si. O espaço dimensional é o simplex de dimensão  $p - 1$ .

Até hoje esse tipo de abordagem ainda é defendida por alguns pesquisadores como por Baxter *et al.* (2005), que compararam essa abordagem com a razão log e concluem que para

algumas análises essa padronização é suficiente e apresenta melhores resultados.

## 2.2 Exclusão de uma Variável (Exc)

Para minimizar o problema da soma constante, podemos escolher retirar uma das variáveis da análise. Essa retirada ocasiona outro problema que é a escolha da variável a ser excluída já que ela não fará parte das análises finais. Também com isso a correlação entre as variáveis restantes seria igual, pois a simples remoção de uma variável não afetaria a correlação entre as outras variáveis e sim somente resolveria o problema da soma constante das variáveis, não solucionando o problema da correlação espúria (Van den Boogaart e Tolosana-Delgado, 2013). O espaço dimensional passa a ser o espaço euclidiano, de dimensão  $p - 1$ .

A variável que pode ser escolhida para ser excluída depende muito da natureza dos dados e deve ser determinada junto com o pesquisador. Um dos critérios adotados pode ser o grau de certeza da medição dos elementos químicos, por exemplo, o elemento que possuir a menor certeza pode ser excluído.

## 2.3 Variável de Referência (Ref)

Para não se retirar uma variável da análise, mas mesmo assim retirar a restrição de soma constante, surgiu a solução de se escolher uma das variáveis como referência e dividir as outras variáveis por essa referência.

Considerando  $x_p$  como referência, a matriz de dados passaria a ser

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{p-1}),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{z}_i$  é  $z_{ki} = \frac{x_{ki}}{x_{kp}}$ , com  $i = 1, 2, \dots, p - 1$ .

Temos que a inversa dessa transformação é dada da seguinte forma:

$$\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{w}_i$  é  $w_{ki} = z_{ki} \cdot x_{ki}$ , com  $i = 1, 2, \dots, p - 1$  e  $w_{kp} = x_{kp}$ .

Nessa metodologia temos que selecionar a variável que será a referência, usualmente escolhida como uma variável que se tem em todas, ou na maioria, das observações para não se ter problemas com divisão por 0, e também a variável com maiores concentrações e certezas de suas medições, ou seja, aquela sobre a qual o pesquisador já tenha conhecimento prévio de que não varia muito de amostra a amostra (Aitchison *et al.*, 2002).

## 2.4 Transformação Log (Log)

A transformação Log é amplamente utilizada na estatística para auxiliar a diminuir a variância dos dados, para deixar os dados mais simétricos e consequentemente aproximar os dados de uma distribuição normal. Mais informações sobre o uso da transformação Log podem ser encontradas em McDonald (2009) e Feng *et al.* (2014).

Essa transformação foi uma das primeiras a ser adotada e sugerida para esse tipo de dados, uma vez que ela facilita a interpretação dos resultados, sendo uma função com inversa conhecida ( $e^x$ ) e também devido ao seu amplo uso e conhecimento.

Nessa transformação são calculados o logaritmo neperiano de todas as variáveis do vetor.

Logo,

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_p),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{z}_i$  é  $z_{ki} = \log(x_{ki})$ , com  $i = 1, 2, \dots, p$ .

Temos que a inversa é dada da seguinte forma:

$$\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{w}_i$  é  $w_{ki} = \exp(z_{ki})$ , com  $i = 1, 2, \dots, p$ .

Nessa transformação pode-se ter problemas com o cálculo de  $\log(0)$ .

## 2.5 Additive Log-Ratio (Alr)

A transformação Alr *Additive log-ratio* (razão logarítmica aditiva) surgiu como uma mistura das transformações descritas nas Seções 2.3 e 2.4. Escolhemos uma variável de referência, digamos  $X_p$ , para ser o denominador da razão e posteriormente se calcula o logaritmo neperiano dessa razão.

Assim,

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{p-1}),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{z}_i$  é  $z_{ki} = \log(\frac{x_{ki}}{x_{kp}})$ , com  $i = 1, 2, \dots, p - 1$ .

Temos que a inversa da transformação é dada da seguinte forma:

$$\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{p-1}),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{w}_i$  é  $w_{ki} = \exp(z_{ki}) \cdot x_{kp}$ , com  $i = 1, 2, \dots, p - 1$  e

$$w_{kp} = x_{kp}.$$

Com isso reduzimos a dimensão do espaço para  $p - 1$ , saindo do espaço simplex para o espaço dos números reais. Assim, como na transformação descrita na Seção 2.3, facilmente conseguimos inverter a função para retornar às variáveis originais.

Devido ao fato de termos que escolher uma variável como referência, temos grande quantidade de transformações possíveis que dependem da escolha da variável referência, mas como demonstrado em Aitchison (1985), independentemente da variável escolhida os resultados da análise serão os mesmos.

Ainda assim temos o problema de possíveis divisões por 0 e também temos que a transformação Alr leva a dados assimétricos (Aitchison *et al.*, 2002).

## 2.6 *Center Log-Ratio* (Clr)

A transformação Clr *Center log-ratio* (razão logarítmica central) aparece como um próximo passo da Alr. Nesta transformação utiliza-se a média geométrica no lugar da variável de referência, evitando, assim, o problema de escolher uma variável além de levar a variáveis simétricas. Assim como a transformação Alr saímos do espaço amostral simplex e passamos a trabalhar no espaço amostral dos números reais (Aitchison *et al.*, 2002).

Temos que:

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_p),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{z}_i$  é  $z_{ki} = \log(\frac{x_{ki}}{\tilde{x}_i})$ , com  $\tilde{x}_i$  sendo a média geométrica da variável  $i$ , com  $i = 1, 2, \dots, p$ .

Temos que a inversa da transformação é dada da seguinte forma:

$$\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p).$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{w}_i$  é  $w_{ki} = \exp(z_{ki}) \cdot \tilde{x}_i$ , com  $i = 1, 2, \dots, p$ .

Nesta transformação também é possível ter problemas de divisão por zero.

## 2.7 Transformação Esférica (Esf)

A transformação esférica foi sugerida por Wang *et al.* (2007), devido à dificuldade de aplicar as transformações descritas nas Seções 2.5 e 2.6 a banco de dados que possuem muitas observações com os valores 0.

Essa transformação mapeia os dados compositivos em uma hiper esfera de dimensão reduzida em um grau de liberdade, devido à soma constante dos dados compositivos. O algoritmo para essa transformação é realizado da seguinte maneira:

Seja

$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{p-1}),$$

em que o  $k$ -ésimo elemento do vetor  $\mathbf{y}_i$  é

$$\begin{aligned} y_{k1} &= \arccos(\sqrt{x_{k1}}), \\ y_{k2} &= \arccos\left(\frac{\sqrt{x_{k2}}}{\sin(y_{k1})}\right), \\ y_{k3} &= \arccos\left(\frac{\sqrt{x_{k3}}}{\sin(y_{k1}) \odot \sin(y_{k2})}\right), \\ &\vdots \\ y_{k(p-1)} &= \arccos\left(\frac{\sqrt{x_{k(p-1)}}}{\sin(y_{k1}) \odot \sin(y_{k2}) \odot \cdots \odot \sin(y_{k(p-2)})}\right). \end{aligned}$$

Essa transformação é indicada quando há valores não identificados. Um dos problemas dessa transformação é que a sua transformação inversa é difícil matematicamente, por causa de sua estrutura recursiva, complicando as análises dos resultados na escala original dos dados, como podemos ver em Neocleous *et al.* (2011).

Note, ainda, que a variável  $x_p$  não é utilizada e isso novamente leva ao problema de escolher qual das variáveis será descartada.



# Capítulo 3

## Base de Dados e Análises

### 3.1 Critérios

Para podermos comparar os desempenhos das transformações vamos utilizar 3 bancos de dados reais com características de número de variáveis e observações diferentes entre si. Para analisar essas bases vamos utilizar a técnica de componentes principais, e não necessariamente replicar as análises dos artigos originais, para comparar as sete transformações enunciadas. A análise de componentes principais foi escolhida, pois ela é comumente utilizada para selecionar as variáveis que são mais relevantes em conjuntos de dados.

Para facilitar as análises utilizamos alguns dos critérios mais comumente usados para avaliar uma análise de componentes principais, entre eles o percentual da variância explicado por componente principal, a carga de cada variável nas componentes principais, o percentual da variância de cada variável explicado por componente principal e a correlação das variáveis com as componentes principais geradas. Os resultados para cada uma das transformações e bases de dados se encontram no Apêndice A.

Para facilitar as comparações fizemos algumas tabelas e gráficos, uma tabela para comparar a proporção explicada de cada componente principal em cada transformação, tabela com o número de componentes principais necessárias para explicar certa porcentagem da variância e o número de componentes para atingir certa comunalidade, tabelas com as maiores cargas e maiores correlações entre as componentes principais e as variáveis originais e também um gráfico com o quanto a carga absoluta de cada variável representa na carga total da primeira componente principal. Para essas construções utilizamos o *software* estatístico R, para mais informações ver Gentleman *et al.* (1997) e Van den Boogaart e Tolosana-Delgado (2013).

## 3.2 Base 1

A primeira base de dados que utilizamos é proveniente de amostras de sedimentos da cordilheira marinha central e sudeste do Oceano Índico. Ela foi utilizada em Kuhn *et al.* (2000) com o intuito de comparar a composição dos sedimentos nessas diferentes regiões. São 161 amostras de sedimentos que são compostas pela porcentagem dos seguintes elementos Silício (Si), Titânio (Ti), Alumínio (Al), Ferro (Fe), Magnésio (Mg), Potássio (K) e Manganês (Mn).

**Tabela 3.1:** Proporção da variância total explicada por componente principal

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	49	55	54	49	53	53	54
CP2	71	75	77	72	76	75	80
CP3	86	91	89	86	89	85	91
CP4	93	98	96	93	96	93	97
CP5	99	100	99	98	99	99	99
CP6	100	100	100	100	100	100	100
CP7	100			100		100	

Podemos notar pela Tabela 3.1 que as transformações que possuem uma variável a menos (Exc, Ref, Alr e Esf) tendem a ter uma explicação maior da variância na primeira componente. Isso provavelmente se deve ao fato de ter que explicar a variância de uma variável a menos. Notamos também que a primeira componente principal da transformação Clr explica, aproximadamente, a mesma proporção da variância do que as transformações que possuem uma variável a menos e que as transformações com 6 variáveis explicam aproximadamente a mesma proporção da variância acumulada na segunda componente principal 75%, menos a transformação Esf que chega a explicar 80% da variância.

**Tabela 3.2:** Números de componentes principais necessárias para explicar certa porcentagem da variância dos dados

	Variância total explicada				Comunalidade
	60%	70%	80%	90%	
Trad	2	2	3	4	3
Exc	2	2	3	3	3
Ref	2	2	3	4	2
Log	2	2	3	4	3
Alr	2	2	3	4	2
Clr	2	2	3	4	2
Esf	2	2	2	3	2

Uma outra visão desta mesma informação se encontra na Tabela 3.2 a qual mostra que as componentes principais explicam a variância aproximadamente na mesma velocidade, pois a quantidade de componentes principais necessárias para explicar as proporções, fixadas, da variância acumulada são praticamente as mesmas.

A exceção fica para a transformação Esf, que necessita de somente 2 e 3 componentes para explicar 80% e 90%, respectivamente, da variância total.

Com relação à comunalidade, a última coluna da Tabela 3.2 mostra que com as transformações Trad, Exc e Log explica-se pelo menos 50% das variâncias individuais de todas as variáveis com três componentes principais. Para as outras transformações esse número é alcançado com duas componentes.

**Tabela 3.3:** *Cargas nas três primeiras componentes principais*

Primeira componente principal							
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Fe	0,52	0,52	0,51	0,52	0,52	0,48	0,54
Si	-0,53	-0,54		-0,51		-0,44	-0,49
Mn	0,49	0,50	0,51	0,50	0,52	0,49	0,51
Mg	0,40	0,41	0,51	0,40	0,50		
Al			0,39		0,37	-0,32	-0,44
K							
Ti						-0,38	
Segunda componente principal							
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Fe							
Si			-0,72		-0,70		
Mn							
Mg						-0,45	-0,63
Al	0,70	0,79	-0,31	0,69	-0,39	-0,48	
K		0,49	-0,56		-0,54	0,66	0,74
Ti	0,62			0,60		-0,33	
Terceira componente principal							
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Fe							
Si			-0,33		0,31		
Mn							
Mg						0,80	0,67
Al		-0,48	-0,48		0,44		-0,58
K	-0,96	0,84	0,77	-0,95	-0,80	0,35	0,40
Ti						-0,32	

**Tabela 3.4:** Correlações entre as variáveis e as três primeiras componentes principais

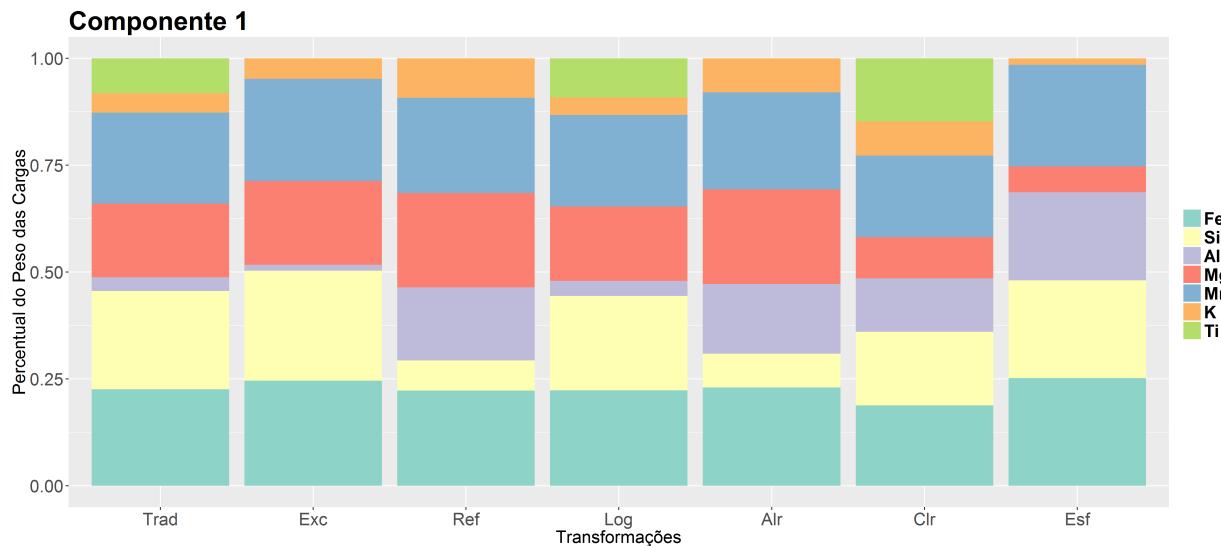
	Primeira componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Fe	0,96	0,94	0,92	0,96	0,93	0,92	0,97
Si	-0,98	-0,99		-0,95		-0,85	-0,88
Mn	0,91	0,92	0,92	0,92	0,92	0,94	0,92
Mg	0,73	0,75	0,91	0,74	0,90		
Al				0,71		0,66	-0,61
K							
Ti						-0,73	
	Segunda componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Fe							
Si				-0,84		-0,83	
Mn							
Mg						-0,55	-0,78
Al	0,88	0,85	-0,36	0,87	-0,47	-0,59	
K		0,52	-0,66		-0,64	0,82	0,91
Ti	0,79			0,76		-0,40	
	Terceira componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Fe							
Si				-0,28		0,27	
Mn							
Mg						0,67	0,56
Al		-0,47	-0,40		0,38		-0,48
K	-0,96	0,83	0,64	-0,93	-0,69	0,30	0,34
Ti						-0,27	

As Tabelas 3.3 e 3.4 e a Figura 3.1 nos mostram a importância de cada variável dentro da primeira componente principal. Foram apresentadas nas tabelas as três primeiras componentes principais porque explicam pelo menos 80% da variância total em qualquer das transformações. Para tornar a visualização mais fácil nas Tabelas 3.3 e 3.4 foram incluídas somente as cargas maiores ou iguais a 0,30 e/ou correlações maiores ou iguais a 0,50, em módulo.

Notamos que as variáveis Fe e Mn são importantes na primeira componente principal em todas as transformações. As variáveis Si, Mg e Al se alternam. A variável Ti aparece com carga considerável nas três primeiras componentes quando a transformação Clr é aplicada e aparece somente na segunda componente principal para Trad e Log. Nas outras transformações ela naturalmente não aparece, pois é considerada como referência. A variável K não aparece na primeira componente em nenhuma das transformações.

As análises efetuadas com as transformações Trad e Log são similares, como também o são as análises com Ref e Alr. Isso nos faz concluir que a aplicação do logaritmo não faz

diferença para a análise.



**Figura 3.1:** Percentual da carga da primeira componente principal explicada por cada variável

Ainda, podemos notar pelos gráficos do Apêndice A que as coordenadas das primeiras e segundas componentes principais das transformações Trad, Exc e Log são similares (Figuras A.2, A.4 e A.8), das transformações Ref e Alr (Figuras A.6 e A.10) também são similares e ambas utilizam a variável Ti como referência, uma na escala normal (Ref) e outra na escala log (Alr) e das transformações Clr e Esf (Figuras A.12 e A.14)

Para essa primeira base de análise notamos que as 7 transformações apresentam resultados similares no percentual da variância explicada para cada componente principal, mas com algumas diferenças das variáveis nas componentes principais e também que as transformações Trad, Exc e Log e as transformações Ref e Alr apresentam coordenadas estimadas muito próximas.

### 3.3 Base 2

A base de dados foi compilada por Grantham e Velbel (1988) que queriam testar como as variáveis presentes nas amostras dependem de certas condições fisiográficas e climáticas por meio da localização onde foi coletado cada sedimento.

Esta base de dados é composta quatro tipos de petrografias. São elas: tipo de sedimentos de rocha poli mineral (Rf), cristais de um único quartzo (Qm), contendo vários quartzos, mas só quartzos (Qp) e mica (M), em porcentagem.

A base é composta de amostras de 8 regiões de 3 tipos de areia grossa, média e fina, sendo que para cada região e tipo de areia foram coletadas 3 amostras, totalizando 72 medições.

**Tabela 3.5:** Proporção da variância total explicada por componente principal

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	85	90	96	79	79	82	82
CP2	93	99	99	92	98	94	95
CP3	100	100	100	96	100	100	100
CP4	100			100		100	

**Tabela 3.6:** Números de componentes principais necessárias para explicar certa porcentagem da variância dos dados

	Variância total explicada				Comunalidade
	60%	70%	80%	90%	70%
Trad	1	1	1	2	1
Exc	1	1	1	1	1
Ref	1	1	1	1	1
Log	1	1	2	2	2
Alr	1	1	2	2	2
Clr	1	1	1	2	1
Esf	1	1	1	2	1

Pela Tabela 3.5 notamos que a primeira componente principal possui um alto nível de explicação e isso é devido ao número reduzido de variáveis utilizadas, somente 4. Notamos que a transformação Alr apesar de ter uma variável a menos, possui a menor explicação juntamente com a transformação Log.

Na Tabela 3.6 notamos que as transformações possuem praticamente a mesma quantidade de componentes principais pra explicar a mesma porcentagem da variância. Notamos somente que as transformações Log e Alr necessitam de duas componentes para explicar 80% da variância enquanto as outras ainda necessitam de uma.

Com relação à comunalidade, a última coluna da Tabela 3.6 mostra que com as transformações Log e Alr explica-se pelo menos 70% das variâncias individuais de todas as variáveis com duas componentes principais. Para as outras transformações essa porcentagem é alcançada com apenas uma componente.

Nas Tabelas 3.7 e 3.8 foram incluídas as duas primeiras componentes principais, porque explicam pelo menos 80% da variância total em qualquer das transformações. Para tornar a visualização mais fácil, foram apresentadas apenas as cargas maiores ou iguais a 0,40 e/ou correlações maiores ou iguais a 0,50 , em módulo. Nessas tabelas e na Figura 3.2 nota-se que todas as variáveis são importantes na primeira componente principal. A variável M aparece somente nas transformações Trad, Log e Clr pois nas demais é considerada como referência.

Percebe-se ainda que as análises com Trad, Log e Clr são similares, sendo que a maior diferença está em que a segunda componente principal tem considerável correlação com M na transformação Log e com Qm na transformação Clr.

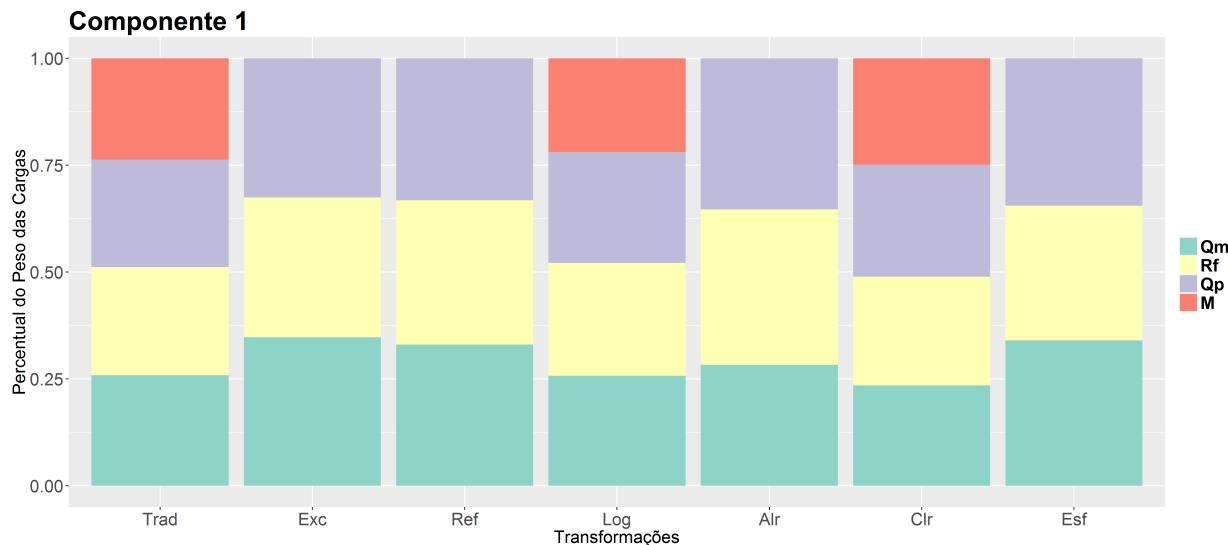
Comentário similar pode ser feito com relação a Ref e Alr e com relação a Exc e Esf. No primeiro caso há correlação considerável entre a segunda componente principal e Qm na transformação Alr e no segundo com Rf na transformação Esf.

**Tabela 3.7:** *Cargas nas duas primeiras componentes principais*

Primeira componente principal							
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Qm	0,52	0,60	-0,57	0,51	0,49	0,47	0,59
Rf	-0,51	-0,57	-0,58	-0,53	0,63	-0,51	-0,55
Qp	-0,50	-0,56	-0,58	-0,52	0,61	-0,52	-0,60
M	0,47			0,44		0,50	
Segunda componente principal							
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Qm	0,52		0,76	0,40	0,87	0,75	0,46
Rf		-0,70					0,83
Qp		0,72	-0,64		-0,41		
M	-0,83			-0,88		-0,60	

**Tabela 3.8:** *Correlações entre as variáveis e as duas primeiras componentes principais*

Primeira componente principal							
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Qm	0,95	0,99	-0,97	0,91	0,75	0,85	0,92
Rf	-0,93	-0,93	-0,99	-0,94	0,97	-0,92	-0,85
Qp	-0,93	-0,93	-0,98	-0,92	0,94	-0,95	-0,94
M	0,87			0,78		0,90	
Segunda componente principal							
	Trad	Exc	Ref	Log	Alr	Clr	Esf
Qm	0,30		0,23	0,28	0,66	0,52	0,29
Rf		-0,36					0,52
Qp		0,37	-0,20		-0,31		
M	-0,48			-0,62		-0,41	



**Figura 3.2:** Percentual da carga da primeira componente principal explicada por cada variável

Podemos notar pelas figuras do Apêndice A que as coordenadas das primeiras e segundas componentes principais das transformações Trad, Log e Clr são semelhantes na distribuição das variáveis dentro do círculo (Figuras A.16, A.22 e A.26). Já as outras transformações podem ser agrupadas em Exc e Esf (Figuras A.18 e A.28) e Ref e Alr (Figuras A.20 e A.24).

## 3.4 Base 3

Esta base de dados, que foi coletada por Tolosana-Delgado e von Eynatten (2010), contém 87 sedimentos glaciais e é composta pelos seguintes óxidos:  $SiO_2$ ,  $TiO_2$ ,  $Al_2O_3$ ,  $MnO$ ,  $MgO$ ,  $CaO$ ,  $Na_2O$ ,  $K_2O$ ,  $P_2O_5$ ,  $Fe_2O_3$  e outras substâncias que não foram capazes de ser medidas, representadas nas tabelas pela palavra *Outros*.

**Tabela 3.9:** Proporção da variância total explicada por componente principal

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	63	61	46	63	51	56	49
CP2	79	78	69	79	73	78	70
CP3	92	92	85	92	87	92	84
CP4	97	97	93	96	95	95	91
CP5	98	99	96	98	98	97	95
CP6	99	99	98	99	99	99	98
CP7	100	100	99	99	99	100	99
CP8	100	100	100	100	100	100	99
CP9	100	100	100	100	100	100	100
CP10	100	100	100	100	100	100	100
CP11	100			100		100	

Pela Tabela 3.9 notamos algumas diferenças expressivas entre os resultados das dife-

rentes transformações no quesito de variância explicada. As transformações Ref, Alr e Esf apresentam a pior explicação para a primeira componente principal apesar de possuírem uma variável a menos. As transformações Trad e Log apresentam as melhores explicações para a primeira componente principal.

**Tabela 3.10:** Números de componentes principais necessárias para explicar certa porcentagem da variância dos dados

	Variância total explicada				Comunalidade
	60%	70%	80%	90%	
Trad	1	2	3	3	2
Exc	1	2	3	3	3
Ref	2	3	3	4	4
Log	1	2	3	3	3
Alr	2	2	3	4	4
Clr	2	2	3	3	2
Esf	2	3	3	4	3

Na Tabela 3.10 notamos que as transformações chegam a percentuais da variância explicada com quantidade de componentes principais diferentes, sendo que as transformações Trad, Exc e Log são as que precisam de menos componentes para explicar a variância.

**Tabela 3.11:** *Cargas nas três primeiras componentes principais*

	Primeira componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
<i>SiO</i> <sub>2</sub>	0,36	0,39	-0,42	0,35	-0,41	-0,35	-0,35
<i>Al</i> <sub>2</sub> O <sub>3</sub>	-0,30	-0,33	-0,45	-0,28	-0,42	-0,38	-0,42
<i>Na</i> <sub>2</sub> O			-0,45		-0,43	-0,39	-0,43
<i>K</i> <sub>2</sub> O	-0,27	-0,29	-0,41		-0,39	-0,32	-0,35
<i>Fe</i> <sub>2</sub> O <sub>3</sub>	-0,36	-0,38		-0,36		0,35	
<i>Ca</i> O			-0,37	-0,28	-0,40		-0,42
<i>Mg</i> O	-0,35	-0,36		-0,35		0,36	0,36
<i>Ti</i> O <sub>2</sub>	-0,35	-0,37		-0,34			
<i>P</i> <sub>2</sub> O <sub>5</sub>				-0,31			
<i>Mn</i> O	-0,35			-0,35		0,32	
<i>Outros</i>	-0,28	-0,29					
	Segunda componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
<i>SiO</i> <sub>2</sub>							
<i>Al</i> <sub>2</sub> O <sub>3</sub>				-0,49			
<i>Na</i> <sub>2</sub> O	-0,51	0,47		-0,61			
<i>K</i> <sub>2</sub> O							
<i>Fe</i> <sub>2</sub> O <sub>3</sub>			0,41		0,55		-0,54
<i>Ca</i> O	-0,54	0,55				0,56	
<i>Mg</i> O					0,45		
<i>Ti</i> O <sub>2</sub>			0,50			0,41	-0,40
<i>P</i> <sub>2</sub> O <sub>5</sub>	-0,42	0,44	0,56		0,47	0,52	0,49
<i>Mn</i> O							
<i>Outros</i>					-0,42		
	Terceira componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
<i>SiO</i> <sub>2</sub>							
<i>Al</i> <sub>2</sub> O <sub>3</sub>	-0,40	-0,42					
<i>Na</i> <sub>2</sub> O	-0,51	-0,60					
<i>K</i> <sub>2</sub> O	-0,45	-0,41		-0,42			
<i>Fe</i> <sub>2</sub> O <sub>3</sub>			-0,57				0,44
<i>Ca</i> O				0,56			
<i>Mg</i> O			-0,57				
<i>Ti</i> O <sub>2</sub>					0,41		-0,47
<i>P</i> <sub>2</sub> O <sub>5</sub>				0,43			0,41
<i>Mn</i> O						-0,45	
<i>Outros</i>					0,57	-0,61	

**Tabela 3.12:** Correlações entre as variáveis e as três primeiras componentes principais

	Primeira componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
<i>SiO<sub>2</sub></i>	0,95	0,96	-0,90	0,92	-0,93	-0,87	-0,77
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,79	-0,82	-0,97	-0,74	-0,95	-0,94	-0,92
<i>Na<sub>2</sub>O</i>			-0,97		-0,97	-0,97	-0,96
<i>K<sub>2</sub>O</i>	-0,72	-0,72	-0,89		-0,87	-0,80	-0,78
<i>Fe<sub>2</sub>O<sub>3</sub></i>	-0,95	-0,93		-0,96		0,87	
<i>CaO</i>			-0,79	-0,73	-0,90		-0,93
<i>MgO</i>	-0,91	-0,89		-0,94		0,90	0,80
<i>TiO<sub>2</sub></i>	-0,91	-0,91		-0,90			
<i>P<sub>2</sub>O<sub>5</sub></i>				-0,82			
<i>MnO</i>	-0,93			-0,93		0,79	
<i>Outros</i>	-0,73	-0,72					
	Segunda componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
<i>SiO<sub>2</sub></i>							
<i>Al<sub>2</sub>O<sub>3</sub></i>				-0,65			
<i>Na<sub>2</sub>O</i>	-0,69	0,61		-0,81			
<i>K<sub>2</sub>O</i>							
<i>Fe<sub>2</sub>O<sub>3</sub></i>			0,63		0,81		-0,79
<i>CaO</i>	-0,72	0,73				0,87	
<i>MgO</i>					0,66		
<i>TiO<sub>2</sub></i>			0,77			0,63	-0,58
<i>P<sub>2</sub>O<sub>5</sub></i>	-0,56	-0,58	0,85		0,70	0,80	0,72
<i>MnO</i>							
<i>Outros</i>					0,66		
	Terceira componente principal						
	Trad	Exc	Ref	Log	Alr	Clr	Esf
<i>SiO<sub>2</sub></i>							
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,48	-0,49					
<i>Na<sub>2</sub>O</i>	-0,61	-0,71					
<i>K<sub>2</sub>O</i>	-0,54	-0,48		-0,49			
<i>Fe<sub>2</sub>O<sub>3</sub></i>			-0,70			0,53	
<i>CaO</i>				0,66			
<i>MgO</i>			-0,70				
<i>TiO<sub>2</sub></i>					0,49		-0,57
<i>P<sub>2</sub>O<sub>5</sub></i>				0,50		0,49	
<i>MnO</i>						-0,55	
<i>Outros</i>					0,69	0,73	

Com relação a comunalidade, Tabela 3.10, mostra-se que com as transformações Trad e Clr explica-se pelo menos 50% das variâncias individuais de todas as variáveis com duas componentes principais. Para as transformações Exc, Log e Esf essa porcentagem é alcançada com três componentes e para Ref e Alr com quatro.

Nas Tabelas 3.11 e 3.12 foram apresentadas as três primeiras componentes principais

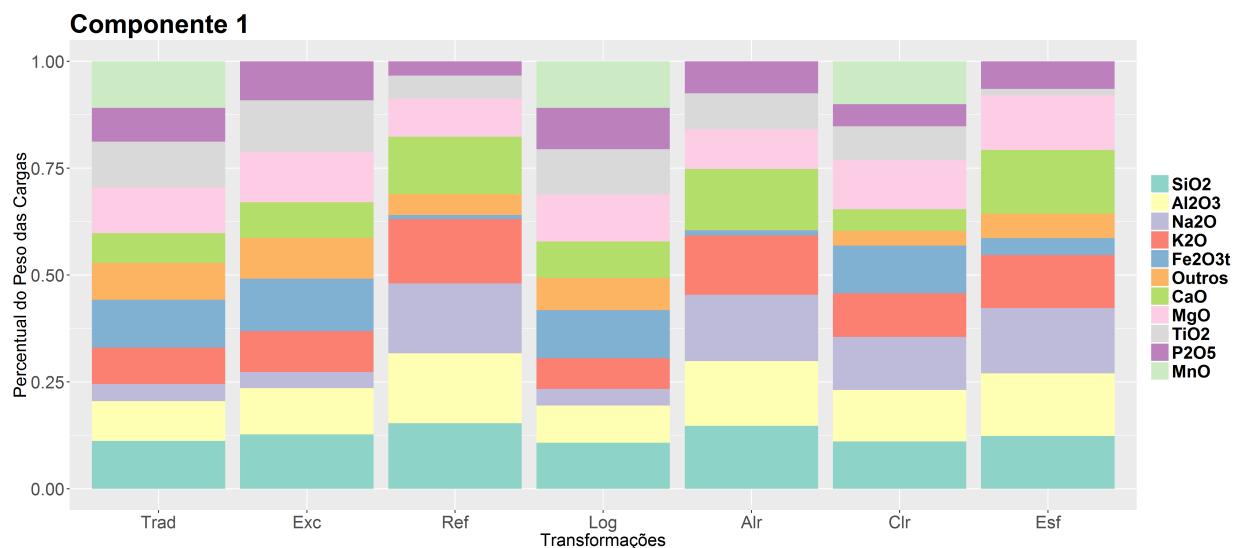
porque explicam pelo menos 70% da variância total em qualquer das transformações. Foram incluídas somente as cargas maiores ou iguais a 0,40 e/ou as correlações maiores ou iguais a 0,70, em módulo.

Essas tabelas e a Figura 3.3 mostram que as análises efetuadas com as transformações Trad, Exc e Log são similares, sendo que para Trad e Exc a primeira componente principal tem correlação considerável com  $K_2O$  e Outros e para Log com  $CaO$  e  $P_2O_5$ . A única diferença entre Trad e Exc é que com Trad a primeira componente tem correlação com  $MnO$  e com Exc não, mas essa variável é aquela que foi excluída quando Exc foi utilizada.

As análises efetuadas com Ref e Alr são também muito similares, o que nos leva a concluir, em conjunto com as observações do parágrafo anterior, que a aplicação do logaritmo não faz diferença na análise dos dados.

A análise com a transformação Esf é bastante parecida com as análises Ref e Alr. A diferença está na terceira componente principal e no fato de que quando a transformação Esf é efetuada, a primeira componente principal tem correlação com a variável  $MgO$ , o que não ocorre com as outras duas transformações.

Notamos pelo gráficos do Apêndice A que as coordenadas das primeiras e segundas componentes principais poderiam ser agrupadas como Trad e Log (Figuras A.30 e A.36 ), Exc e Clr (Figuras A.32 e A.40 ) e Ref, Alr e Esf (Figuras A.34, A.38 e A.42 ).



**Figura 3.3:** Percentual da carga da primeira componente principal explicada por cada variável

# Capítulo 4

## Simulação

Neste capítulo vamos utilizar um estudo de simulações para podermos aplicar as transformações em conjuntos de dados similares e compararmos os resultados encontrados. Devido à possível interferência que o modo de simulação dos dados possa gerar, vamos simular os dados de 3 formas diferentes.

### 4.1 Dirichlet

Escolhemos a distribuição Dirichlet por causa de sua forma. Ela consiste em uma distribuição discreta multivariada com um vetor de parâmetros  $\boldsymbol{\alpha}$  com  $\alpha_i > 0$  e com  $i = 1, 2, \dots, p$ , com  $p \geq 2$ . Sua função de probabilidade está definida da seguinte forma:

$$f(x_1, x_2, \dots, x_p; \alpha_1, \alpha_2, \dots, \alpha_p) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^p x_i^{\alpha_i - 1},$$

em que  $B(\boldsymbol{\alpha})$  é definida em termos da função gama da seguinte forma

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^p \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^p \alpha_i)}.$$

Como a função  $\Gamma(\alpha)$  é definida como:

$$\begin{aligned}\Gamma(\alpha) &= (n-1)! \text{ para } \alpha \text{ inteiro positivo e} \\ \Gamma(\alpha) &= \int_0^\infty v^{\alpha-1} e^{-v} dv, \text{ para } \alpha \text{ real positivo}\end{aligned}$$

é fácil ver que:

$$E[X_i] = \frac{\alpha_i}{\sum_{i=1}^p \alpha_i}.$$

Utilizando a distribuição gama, podemos simular uma distribuição Dirichlet, assim

como demonstrado em Luc (1986). Podemos simular uma estrutura de dados composticionais com  $p$  variáveis utilizando o seguinte algoritmo.

1. Fixa-se  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)$
2. Simulam-se  $n$  valores de  $y_i \sim Gama(\alpha_i, 1)$  com  $i = 1, 2, \dots, p$
3. Calculam-se  $x_i = \frac{y_i}{\sum_{i=1}^p y_i}$  com  $i = 1, 2, \dots, p$

## 4.2 Uniforme

Outro modo de simularmos os dados é por meio de distribuições uniformes independentes. Seja

$$f(x_i) = \begin{cases} \frac{1}{b-a}, & \text{para } a \leq x_i \leq b, \\ 0 & \text{para } x_i < a \text{ ou } x_i > b \end{cases}.$$

Logo, temos que:

$$\begin{aligned} E[X_i] &= \frac{1}{2}(a + b) \\ Var[X_i] &= \frac{1}{12}(b - a)^2 \end{aligned}$$

Podemos fixar os parâmetros da seguinte forma: seja  $E[X_i] = m_i$  e  $Var[X_i] = s_i^2$  com  $a \leq m_i \leq b$ . Temos que  $a = m_i - \sqrt{3}s_i$  e  $b = m_i + \sqrt{3}s_i$ . Com isso podemos utilizar o seguinte algoritmo de aceitação-rejeição para simular um conjunto de dados composticionais.

1. Fixa-se  $\mathbf{m} = (m_1, m_2, \dots, m_{p-1})$ , vetor de médias das  $p - 1$  variáveis
2. Fixa-se  $\mathbf{s}^2 = (s_1^2, s_2^2, \dots, s_{p-1}^2)$ , vetor de variâncias das  $p - 1$  variáveis
3. Simulam-se  $n$  valores de  $x_i \sim Unif(m_i - \sqrt{3}s_i, m_i + \sqrt{3}s_i)$  com  $i = 1, 2, \dots, p - 1$
4. Calcula-se  $x_p = 1 - \sum_{i=1}^{p-1} x_i$
5. Caso algum  $x_p$  seja menor que 0 rejeita-se a observação e simula-se outra

## 4.3 Normal

Por último, também vamos utilizar distribuições normais truncadas entre 0 e 1, definidas da seguinte forma: seja  $x_i \sim N_t(\mu, \sigma^2)$  com  $x_i \in (0, 1)$ . Com isso temos que:

$$f(x_i, \mu, \sigma, 0, 1) = \frac{\frac{1}{\sigma}\phi\left(\frac{x_i - \mu}{\sigma}\right)}{\Phi\left(\frac{1 - \mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)},$$

em que  $\Phi$  denota a função de distribuição acumulada e  $\phi$  a função densidade de probabilidade da distribuição  $N(0, 1)$ .

Para simular um valor da normal truncada vamos utilizar o algoritmo desenvolvido em Geweke (1992). Para simularmos os dados composicionais vamos utilizar um algoritmo semelhante ao utilizado para a distribuição uniforme, da seguinte forma:

1. Fixa-se  $\mathbf{m} = (m_1, m_2, \dots, m_{p-1})$ , vetor de médias das  $p - 1$  variáveis
2. Fixa-se  $\mathbf{s}^2 = (s_1^2, s_2^2, \dots, s_{p-1}^2)$ , vetor de variâncias das  $p - 1$  variáveis
3. Simulam-se  $n$  valores de  $x_i \sim N_t(m_i, s_i^2)$  truncada no intervalo  $[0, 1]$ , com  $i = 1, 2, \dots, p - 1$ .
4. Calcula-se  $x_p = 1 - \sum_{i=1}^{p-1} x_i$
5. Caso algum  $x_p$  seja menor que 0 rejeita-se a observação e simula-se outra

Escolhemos utilizar essa distribuição truncada ao invés da distribuição normal usual devido à maior aceitação no algoritmo de aceitação-rejeição para simular os dados compostionais.

## 4.4 Critérios

Para cada base simulada vamos analisar os seguintes critérios:

- O intervalo de confiança da proporção da variância acumulada explicada por componente principal em cada transformação, ou seja, para cada base simulada vamos calcular a variância acumulada explicada, e depois calcular a média e a variância de todos esses valores. O intervalo de confiança é dado por média  $\pm 1,96$  desvio-padrão. Para facilitar a denotação iremos nos referir a esse critério como explicação da variância.
- Para uma melhor visualização das diferenças entre as transformações criamos dendrogramas da média da diferença absoluta entre as cargas das variáveis em todas as transformações. Calculamos da seguinte forma:
  1. Padroniza-se o sinal das cargas da componente principal, com base no sinal da carga da primeira variável, ou seja, se o sinal for negativo inverte-se o sinal de todas as cargas das variáveis.
  2. Para todas as combinações de transformações, duas a duas, calcula-se a média dos módulos das diferenças entre as cargas das variáveis nas componentes principais, em cada base simulada.

3. Calcula-se a média das diferenças absolutas entre transformações de todas as bases simuladas.
4. Considera-se o valor dessa média como uma medida de distância entre as transformações e desenha-se o dendrograma, com base no método do vizinho mais distante.

Para facilitar a denotação iremos nos referir a esse critério como interpretação das componentes principais.

Para a realização das simulações fixamos em 1000 o número de observações de cada base e também o número de ensaios.

## 4.5 Simulação 1

A primeira estrutura de base simulada possui 4 variáveis com proporções aproximadas entre si, ou seja, com valores em torno de 0,25.

### 4.5.1 Dirichlet

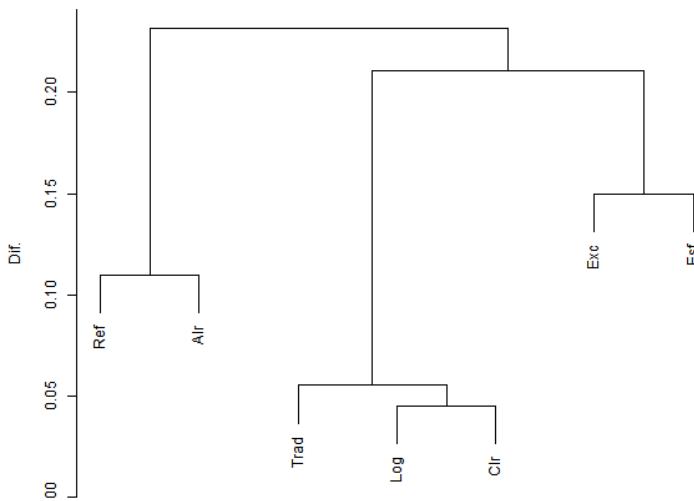
Para simularmos a base, definimos  $\alpha = (1; 1; 1; 1)$ . Podemos notar pela Tabela 4.1 que a transformação Ref possui a melhor explicação da variância, seguida das transformações Alr e Exc. As transformações Trad, Log, Clr e Esf têm explicação da variância baixa nas primeiras componentes principais. Pela Tabela 4.2 e Figura 4.1 notamos que as transformações Log, Clr e Trad levam a interpretações das componentes principais muito similares. Depois as transformações Ref e Alr possuem interpretações das componentes principais similares e por último as transformações Exc e Esf apresentam resultados similares por esse critério.

**Tabela 4.1:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[34,39 ; 34,46]	[45,49 ; 45,57]	[77,22 ; 78,66]	[31,68 ; 31,76]	[66,62 ; 66,82]	[34,81 ; 34,9]	[34,93 ; 35,02]
CP2	[67,75 ; 67,82]	[89,32 ; 89,39]	[93,95 ; 94,56]	[61,91 ; 62,00]	[83,96 ; 84,07]	[68,15 ; 68,24]	[68,24 ; 68,33]
CP3	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[90,68 ; 90,75]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP4	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.2:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,1776	0,2318	0,0503	0,2281	0,0555	0,1773
Exc	0,1776	0,0000	0,2023	0,2006	0,2227	0,2109	0,1499
Ref	0,2318	0,2023	0,0000	0,2281	0,1097	0,2226	0,1774
Log	0,0503	0,2006	0,2281	0,0000	0,2185	0,0452	0,1817
Alr	0,2281	0,2227	0,1097	0,2185	0,0000	0,2001	0,1713
Clr	0,0555	0,2109	0,2226	0,0452	0,2001	0,0000	0,1797
Esf	0,1773	0,1499	0,1774	0,1817	0,1713	0,1797	0,0000



**Figura 4.1:** Dendrogramma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

#### 4.5.2 Uniforme

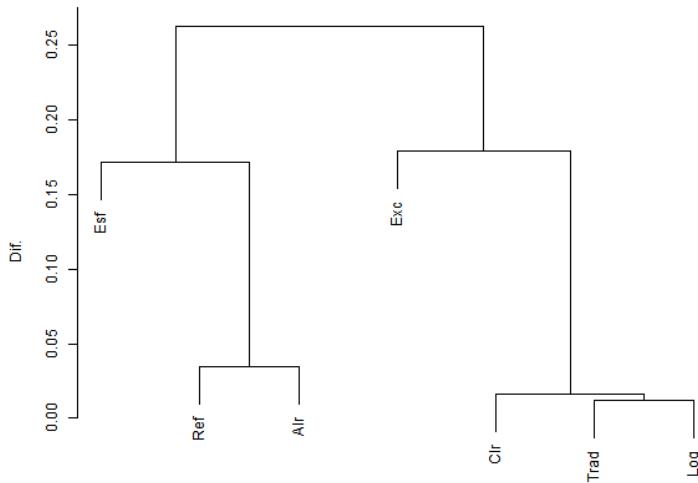
Para simular, fixamos  $\mathbf{m} = (0, 25; 0, 25; 0, 25)$  e  $\mathbf{s}^2 = (0, 008; 0, 008; 0, 008)$ . Podemos notar pela Tabela 4.3, que a transformação Ref possui a maior explicação da variância, seguida das transformações Alr e Exc, que possuem médias muito próximas, depois temos as transformações Clr, Trad e Log e, por último, a Esf. Com base na Tablea 4.4 e na Figura 4.2 notamos no dendrogramma o grupo de transformações Log, Trad e Clr com interpretações das componentes principais similares; o mesmo ocorre para as transformações Ref e Alr. Já as transformações Exc e Esf não se encaixaram em nenhum dos grupos formados tendo interpretações diferentes das outras transformações.

**Tabela 4.3:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[46,11 ; 46,21]	[57,10 ; 57,24]	[62,35 ; 62,47]	[43,02 ; 43,12]	[57,39 ; 57,52]	[51,27 ; 51,46]	[34,73 ; 34,83]
CP2	[73,90 ; 73,98]	[93,00 ; 93,03]	[85,05 ; 85,16]	[70,42 ; 70,51]	[83,43 ; 83,53]	[76,40 ; 76,51]	[68,04 ; 68,14]
CP3	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[96,12 ; 96,19]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP4	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.4:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,1769	0,2622	0,0123	0,2603	0,0163	0,2130
Exc	0,1769	0,0000	0,2497	0,1786	0,2531	0,1789	0,1753
Ref	0,2622	0,2497	0,0000	0,2604	0,0346	0,2623	0,1714
Log	0,0123	0,1786	0,2604	0,0000	0,2594	0,0147	0,2124
Alr	0,2603	0,2531	0,0346	0,2594	0,0000	0,2632	0,1667
Clr	0,0163	0,1789	0,2623	0,0147	0,2632	0,0000	0,2165
Esf	0,2130	0,1753	0,1714	0,2124	0,1667	0,2165	0,0000



**Figura 4.2:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

### 4.5.3 Normal

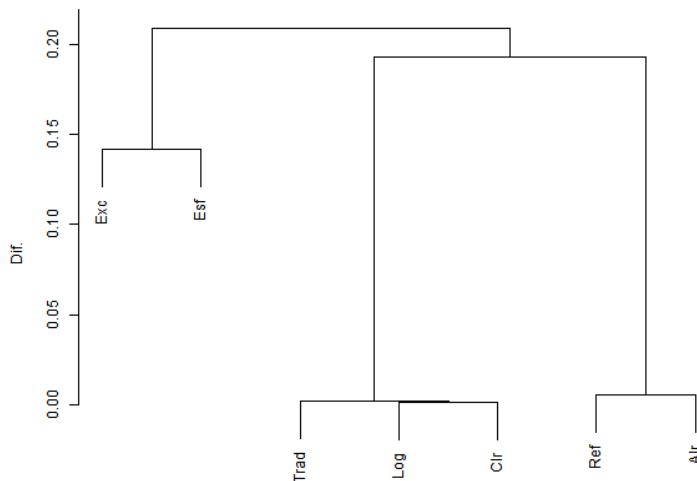
Para simular, fixamos  $\mathbf{m} = (0, 25; 0, 25; 0, 25)$  e  $\mathbf{s}^2 = (0, 008; 0, 008; 0, 008)$ . Podemos notar pela Tabela 4.5, que a transformação Ref e Alr possuem explicações da variância similares e também são as maiores. Já as transformações Exc, Clr, Trad, Log e Esf, têm médias similares e mais baixas. Com base na Tabela 4.6 e na Figura 4.3 notamos no dendrograma que o grupo de transformações Log, Trad e Clr possuem interpretações das componentes principais similares; o mesmo ocorre para as transformações Ref e Alr, seguidas das transformações Exc e Esf.

**Tabela 4.5:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[50,02 ; 50,14]	[51,14 ; 52,64]	[67,48 ; 69,33]	[49,98 ; 50,10]	[67,36 ; 69,21]	[50,07 ; 50,19]	[48,94 ; 50,44]
CP2	[75,81 ; 75,90]	[84,45 ; 85,96]	[86,94 ; 87,61]	[75,77 ; 75,86]	[86,88 ; 87,56]	[75,83 ; 75,91]	[77,58 ; 78,52]
CP3	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP4	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	[100,00 ; 100,00]

**Tabela 4.6:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,1772	0,1918	0,0020	0,1922	0,0016	0,1939
Exc	0,1772	0,0000	0,2088	0,1778	0,2091	0,1775	0,1419
Ref	0,1918	0,2088	0,0000	0,1928	0,0057	0,1924	0,1595
Log	0,0020	0,1778	0,1928	0,0000	0,1922	0,0015	0,1944
Alr	0,1922	0,2091	0,0057	0,1922	0,0000	0,1921	0,1598
Clr	0,0016	0,1775	0,1924	0,0015	0,1921	0,0000	0,1942
Esf	0,1939	0,1419	0,1595	0,1944	0,1598	0,1942	0,0000



**Figura 4.3:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

## 4.6 Simulação 2

A próxima estrutura de base simulada possui uma das variáveis com proporção maior e as outras 3 com proporções similares entre si, ou seja, a proporção média da primeira variável será 3 vezes a proporção das outras..

### 4.6.1 Dirichlet

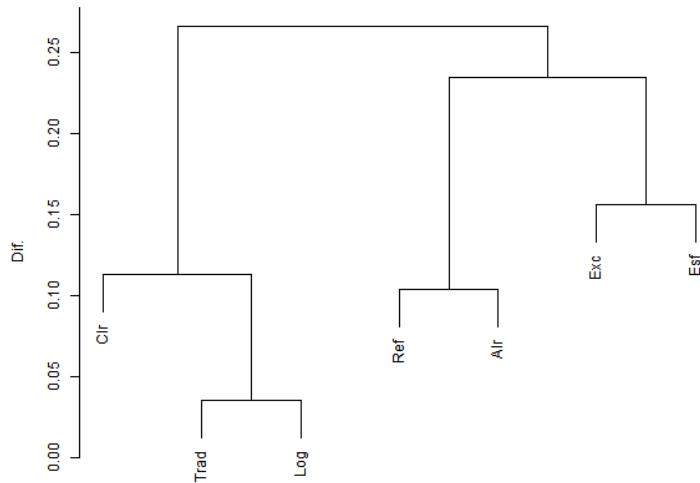
Para simularmos a base, definimos  $\alpha = (3; 1; 1; 1)$ . Pela Tabela 4.7, notamos que a transformação Ref obtém resultados superiores no quesito de explicação da variância seguida da transformação Alr. As demais Exc, Trad, Log, Clr e Esf apresentam explicações similares. Na Tabela 4.8 e na Figura 4.4 notamos que as transformações Trad e Log têm interpretações das componentes principais parecidas, com a transformação Clr próxima delas também. As transformações Ref e Alr também levam a interpretação similar, seguidas da Exc e Esf.

**Tabela 4.7:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[40,05 ; 40,13]	[51,56 ; 51,68]	[80,77 ; 82,03]	[36,43 ; 36,54]	[72,67 ; 72,85]	[36,56 ; 36,64]	[34,99 ; 35,09]
CP2	[70,66 ; 70,73]	[91,71 ; 91,77]	[95,43 ; 95,91]	[65,31 ; 65,40]	[89,39 ; 89,48]	[71,41 ; 71,47]	[68,30 ; 68,40]
CP3	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[92,49 ; 92,56]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP4	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.8:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,1714	0,2641	0,0354	0,2652	0,1132	0,2176
Exc	0,1714	0,0000	0,2195	0,1761	0,2343	0,2520	0,1559
Ref	0,2641	0,2195	0,0000	0,2625	0,1036	0,2229	0,1854
Log	0,0354	0,1761	0,2625	0,0000	0,2664	0,1111	0,2124
Alr	0,2652	0,2343	0,1036	0,2664	0,0000	0,1802	0,1941
Clr	0,1132	0,2520	0,2229	0,1111	0,1802	0,0000	0,2155
Esf	0,2176	0,1559	0,1854	0,2124	0,1941	0,2155	0,0000

**Figura 4.4:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

#### 4.6.2 Uniforme

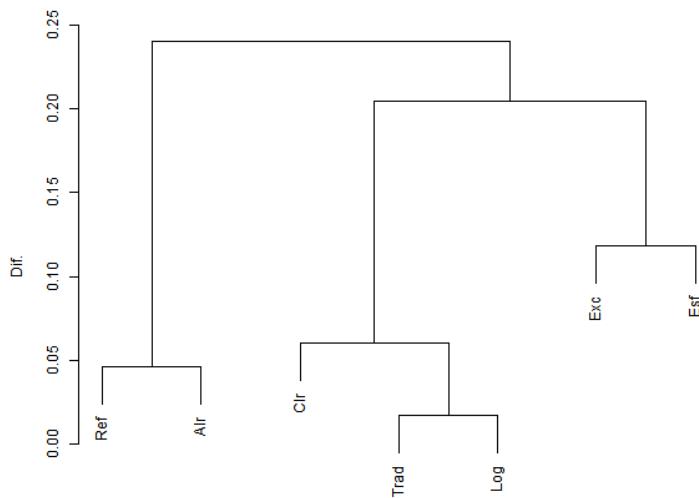
Para simular, fixamos  $\mathbf{m} = (0,5; 0,1667; 0,1667)$  e mantivemos a variância de  $s^2 = (0,008; 0,008; 0,008)$ . Pela Tabela 4.9, vemos que a transformação Ref possui a maior variância explicada, seguida de perto pela transformação Alr, depois temos a transformação Exc, por último, temos as transformações Trad , Log, Clr e Esf, que possuem médias parecidas. Pela Tabela 4.10 e na Figura 4.5 notamos que as transformações Trad e Log possuem interpretações das componentes principais parecidas, seguidas da transformação Clr. Depois temos as transformações Ref e Alr e, por último, Exc e Esf.

**Tabela 4.9:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[42,49 ; 42,58]	[53,73 ; 53,86]	[80,73 ; 80,88]	[39,31 ; 39,41]	[74,14 ; 74,28]	[42,35 ; 42,49]	[41,47 ; 41,59]
CP2	[72,06 ; 72,14]	[92,03 ; 92,07]	[96,01 ; 96,07]	[68,11 ; 68,20]	[93,95 ; 93,99]	[74,34 ; 74,42]	[74,71 ; 74,84]
CP3	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[95,23 ; 95,29]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP4	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.10:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,1577	0,2312	0,0175	0,2298	0,0528	0,2000
Exc	0,1577	0,0000	0,2226	0,1521	0,2213	0,1885	0,1184
Ref	0,2312	0,2226	0,0000	0,2404	0,0459	0,2020	0,1993
Log	0,0175	0,1521	0,2404	0,0000	0,2388	0,0603	0,2047
Alr	0,2298	0,2213	0,0459	0,2388	0,0000	0,2015	0,1974
Clr	0,0528	0,1885	0,2020	0,0603	0,2015	0,0000	0,1769
Esf	0,2000	0,1184	0,1993	0,2047	0,1974	0,1769	0,0000

**Figura 4.5:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

### 4.6.3 Normal

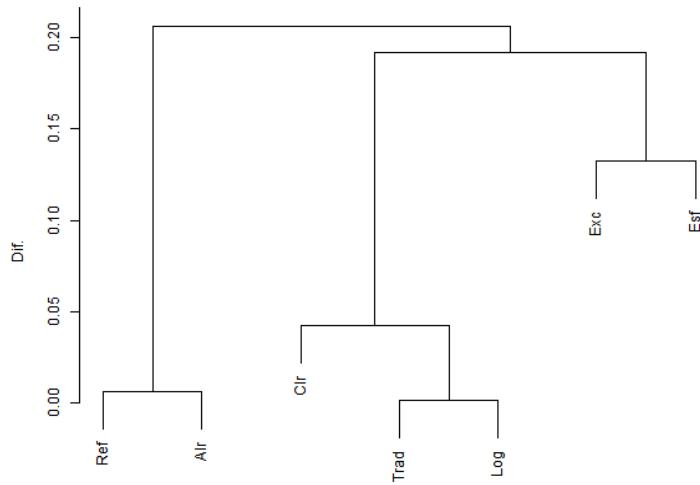
Fixamos  $\mathbf{m} = (0,5; 0,1667; 0,1667)$  e  $\mathbf{s}^2 = (0,008; 0,008; 0,008)$ , para simular os dados normais. Podemos notar pela Tabela 4.11, que a transformação Ref e Alr possuem variância explicada similares e também são as maiores. Na Tabela 4.12 e na Figura 4.6 notamos que as transformações Trad e Log possuem interpretações das componentes principais similares, seguidas da transformação Clr. Depois temos as transformações Ref e Alr e, por último, Exc e Esf.

**Tabela 4.11:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[49,97 ; 50,09]	[47,33 ; 48,91]	[80,12 ; 81,54]	[49,88 ; 50,01]	[79,91 ; 81,33]	[55,39 ; 55,52]	[56,12 ; 56,93]
CP2	[75,80 ; 75,88]	[80,70 ; 82,28]	[95,44 ; 95,74]	[75,72 ; 75,81]	[95,40 ; 95,70]	[79,76 ; 79,86]	[83,15 ; 83,38]
CP3	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP4	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.12:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,1729	0,1932	0,0015	0,1927	0,0425	0,1814
Exc	0,1729	0,0000	0,2065	0,1732	0,2066	0,1921	0,1327
Ref	0,1932	0,2065	0,0000	0,1933	0,0062	0,1393	0,1866
Log	0,0015	0,1732	0,1933	0,0000	0,1928	0,0424	0,1812
Alr	0,1927	0,2066	0,0062	0,1928	0,0000	0,1366	0,1862
Clr	0,0425	0,1921	0,1393	0,0424	0,1366	0,0000	0,1918
Esf	0,1814	0,1327	0,1866	0,1812	0,1862	0,1918	0,0000

**Figura 4.6:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

## 4.7 Simulação 3

Para analisarmos se os resultados mudam quando se tem uma maior quantidade de variáveis, agora vamos simular estruturas de base de dados compostionais com 8 variáveis, primeiramente com as proporções parecidas e depois com uma proporção maior que as demais.

### 4.7.1 Dirichlet

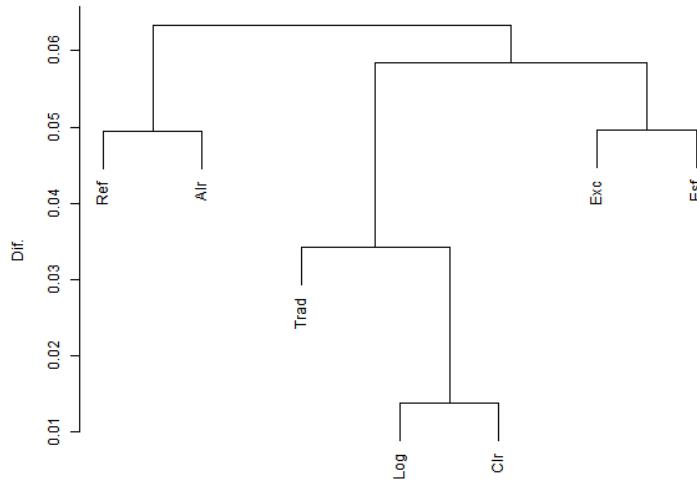
Para simularmos a base, definimos  $\alpha = (1; 1; 1; 1; 1; 1; 1; 1)$ . Pela Tabela 4.13 agora notamos que a transformação Ref tem um resultado melhor no quesito variância explicada, seguida pela transformação Alr; as transformações Trad, Exc, Log, Clr e Esf possuem variâncias explicadas muito baixas. Pela Tabela 4.14 e Figura 4.7, vemos que as transformações Log e Clr possuem interpretações das componentes principais mais próximas, seguidas da transformação Trad. Depois temos as transformações Ref e Alr, e Exc e Esf.

**Tabela 4.13:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[15,65 ; 15,71]	[17,73 ; 17,79]	[71,07 ; 72,47]	[15,17 ; 15,22]	[57,07 ; 57,27]	[15,97 ; 16,01]	[15,98 ; 16,03]
CP2	[30,70 ; 30,76]	[34,76 ; 34,82]	[85,33 ; 86,40]	[29,76 ; 29,83]	[65,11 ; 65,28]	[31,21 ; 31,28]	[31,24 ; 31,31]
CP3	[45,36 ; 45,42]	[51,34 ; 51,41]	[92,24 ; 92,95]	[43,82 ; 43,89]	[72,74 ; 72,87]	[45,95 ; 46,02]	[45,98 ; 46,05]
CP4	[59,64 ; 59,71]	[67,42 ; 67,49]	[95,96 ; 96,40]	[57,38 ; 57,46]	[80,01 ; 80,12]	[60,22 ; 60,29]	[60,24 ; 60,32]
CP5	[73,53 ; 73,60]	[83,07 ; 83,11]	[98,12 ; 98,36]	[70,53 ; 70,60]	[86,99 ; 87,06]	[74,02 ; 74,08]	[74,03 ; 74,09]
CP6	[87,03 ; 87,06]	[98,00 ; 98,00]	[99,37 ; 99,48]	[83,19 ; 83,24]	[93,73 ; 93,78]	[87,26 ; 87,31]	[87,29 ; 87,35]
CP7	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[95,13 ; 95,17]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP8	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.14:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,0470	0,0634	0,0343	0,0597	0,0341	0,0496
Exc	0,0470	0,0000	0,0588	0,0579	0,0582	0,0585	0,0497
Ref	0,0634	0,0588	0,0000	0,0634	0,0495	0,0632	0,0580
Log	0,0343	0,0579	0,0634	0,0000	0,0553	0,0138	0,0512
Alr	0,0597	0,0582	0,0495	0,0553	0,0000	0,0531	0,0532
Clr	0,0341	0,0585	0,0632	0,0138	0,0531	0,0000	0,0512
Esf	0,0496	0,0497	0,0580	0,0512	0,0532	0,0512	0,0000

**Figura 4.7:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

#### 4.7.2 Uniforme

Para simular, fixamos  $\mathbf{m} = (0, 125; 0, 125; 0, 125; 0, 125; 0, 125; 0, 125; 0, 125; 0, 125)$  e  $\mathbf{s}^2 = (0,0008; 0,0008; 0,0008; 0,0008; 0,0008; 0,0008; 0,0008)$ . Podemos notar pela Tabela 4.15 que as transformações Ref e Alr têm as maiores variâncias explicadas, seguidas da transformação Clr e depois temos a Exc e Trad, Log e Esf. Pela Tabela 4.16 e Figura 4.8, vemos que as interpretações das componentes principais obtidas pelas transformações Log, Clr e Trad

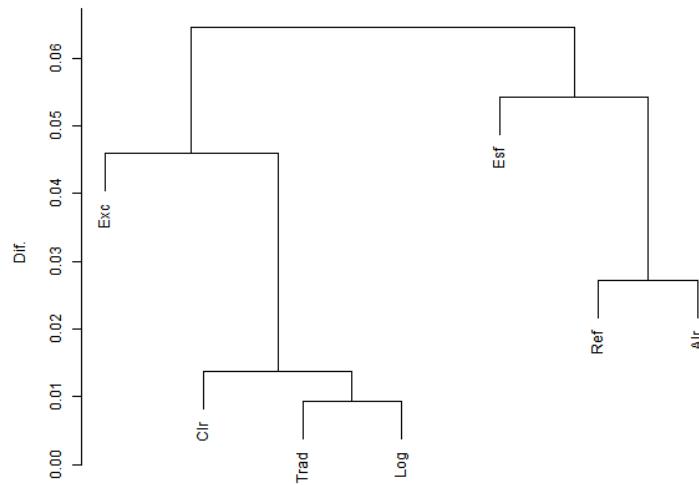
são similares, o mesmo ocorre para as transformações Ref e Alr. Já as transformações Exc e Esf não se enquadram em nenhum grupo.

**Tabela 4.15:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[23,00 ; 23,06]	[25,51 ; 25,58]	[53,19 ; 53,29]	[21,54 ; 21,60]	[51,34 ; 51,45]	[28,64 ; 28,81]	[15,93 ; 15,98]
CP2	[37,20 ; 37,27]	[41,48 ; 41,56]	[65,35 ; 65,45]	[35,69 ; 35,76]	[64,39 ; 64,49]	[41,82 ; 41,95]	[31,12 ; 31,19]
CP3	[50,73 ; 50,80]	[56,73 ; 56,81]	[73,05 ; 73,14]	[49,19 ; 49,26]	[72,22 ; 72,31]	[54,38 ; 54,50]	[45,86 ; 45,94]
CP4	[63,80 ; 63,86]	[71,35 ; 71,42]	[80,29 ; 80,37]	[62,19 ; 62,26]	[79,65 ; 79,72]	[66,47 ; 66,57]	[60,13 ; 60,20]
CP5	[76,33 ; 76,40]	[85,41 ; 85,47]	[87,19 ; 87,24]	[74,80 ; 74,86]	[86,78 ; 86,84]	[78,12 ; 78,19]	[73,95 ; 74,01]
CP6	[88,44 ; 88,50]	[99,00 ; 99,00]	[93,92 ; 93,95]	[86,89 ; 86,94]	[93,60 ; 93,66]	[89,28 ; 89,34]	[87,21 ; 87,27]
CP7	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[98,29 ; 98,34]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP8	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.16:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,0442	0,0616	0,0094	0,0605	0,0138	0,0553
Exc	0,0442	0,0000	0,0636	0,0451	0,0646	0,0460	0,0586
Ref	0,0616	0,0636	0,0000	0,0617	0,0273	0,0628	0,0542
Log	0,0094	0,0451	0,0617	0,0000	0,0599	0,0111	0,0549
Alr	0,0605	0,0646	0,0273	0,0599	0,0000	0,0602	0,0529
Clr	0,0138	0,0460	0,0628	0,0111	0,0602	0,0000	0,0565
Esf	0,0553	0,0586	0,0542	0,0549	0,0529	0,0565	0,0000



**Figura 4.8:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

#### 4.7.3 Normal

Para simular, fixamos  $\mathbf{m} = (0, 125; 0, 125; 0, 125; 0, 125; 0, 125; 0, 125; 0, 125)$  e  $s^2 = (0,0008; 0,0008; 0,0008; 0,0008; 0,0008; 0,0008; 0,0008)$ . Podemos notar pela Tabela 4.17, que as transformações Ref e Alr possuem variâncias explicadas similares e também são as

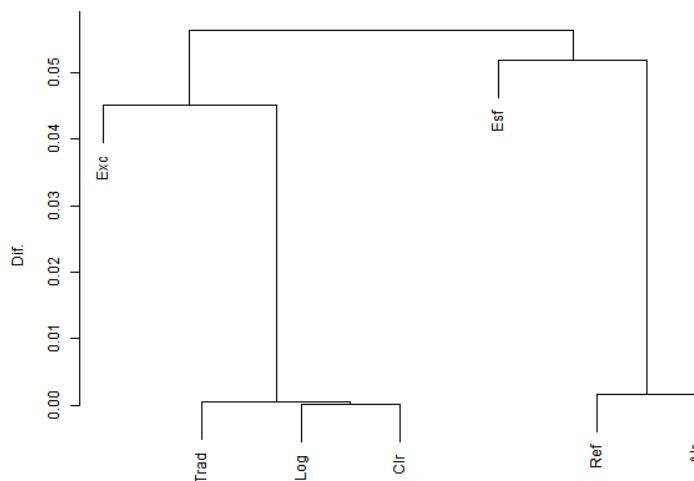
maiores, seguidas da transformação Esf, e, por último, temos as transformações Clr, Exc, Trad, Log que têm médias parecidas e são baixas. Pela Tabela 4.18 e Figura 4.9 notamos que as transformações Trad, Log e Clr levam a interpretações similares das componentes principais. Depois temos as transformações Ref e Alr e, já as transformações Esf e Exc não se enquadram em nenhum grupo.

**Tabela 4.17:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[25,04 ; 25,11]	[23,54 ; 24,22]	[63,25 ; 65,55]	[25,04 ; 25,11]	[63,25 ; 65,55]	[25,05 ; 25,12]	[28,83 ; 30,37]
CP2	[38,86 ; 38,93]	[39,01 ; 39,70]	[72,74 ; 74,40]	[38,86 ; 38,93]	[72,73 ; 74,39]	[38,86 ; 38,93]	[43,11 ; 44,51]
CP3	[52,05 ; 52,12]	[53,83 ; 54,53]	[78,75 ; 80,05]	[52,05 ; 52,12]	[78,75 ; 80,04]	[52,05 ; 52,12]	[56,62 ; 57,86]
CP4	[64,76 ; 64,83]	[68,08 ; 68,78]	[84,44 ; 85,39]	[64,76 ; 64,83]	[84,44 ; 85,39]	[64,76 ; 64,83]	[69,39 ; 70,43]
CP5	[77,02 ; 77,08]	[81,82 ; 82,51]	[89,83 ; 90,43]	[77,02 ; 77,08]	[89,83 ; 90,43]	[77,02 ; 77,08]	[81,22 ; 82,01]
CP6	[88,84 ; 88,89]	[94,91 ; 95,58]	[95,18 ; 95,49]	[88,84 ; 88,89]	[95,18 ; 95,49]	[88,84 ; 88,89]	[91,82 ; 92,30]
CP7	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP8	[100,00 ; 100,00]			[100,00 ; 100,00]			[100,00 ; 100,00]

**Tabela 4.18:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,0451	0,0469	0,0005	0,0467	0,0005	0,0562
Exc	0,0451	0,0000	0,0564	0,0451	0,0563	0,0451	0,0539
Ref	0,0469	0,0564	0,0000	0,0469	0,0017	0,0469	0,0518
Log	0,0005	0,0451	0,0469	0,0000	0,0468	0,0002	0,0562
Alr	0,0467	0,0563	0,0017	0,0468	0,0000	0,0468	0,0518
Clr	0,0005	0,0451	0,0469	0,0002	0,0468	0,0000	0,0562
Esf	0,0562	0,0539	0,0518	0,0562	0,0518	0,0562	0,0000



**Figura 4.9:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

## 4.8 Simulação 4

Agora vamos simular estruturas de base de dados compostionais com 8 variáveis, com a primeira variável tendo uma concentração maior, ou seja, a primeira variável será, em média 3 vezes as outras, assim como na Simulação 2, mas desta vez com 8 variáveis.

### 4.8.1 Dirichlet

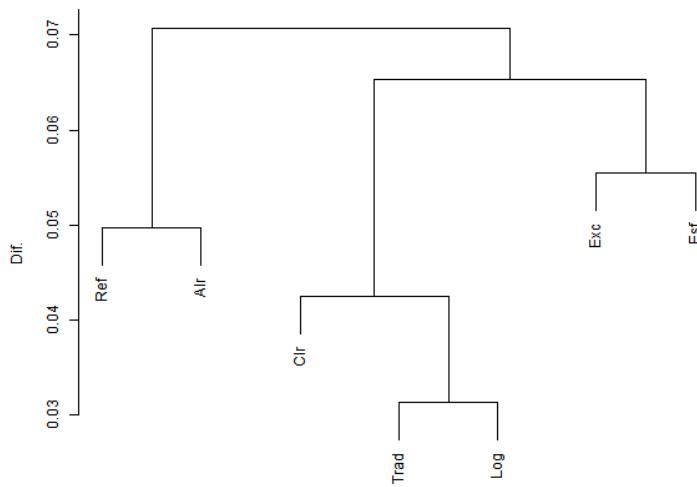
Para simularmos a base, definimos  $\alpha = (3, 1, 1, 1, 1, 1, 1, 1)$ . Pela Tabela 4.19 agora notamos que a transformação Ref tem o melhor resultado no quesito de variância explicada, a transformação Alr possui uma variância explicada levemente menor, seguidas da transformação Exc, e por último temos as transformações Trad, Log, Clr e Esf que têm médias parecidas e são baixas. Pela Tabela 4.20 e Figura 4.10 notamos que as transformações Trad e Log possuem interpretações das componentes principais similares, seguidas da transformação Clr. Depois temos as transformações Ref e Alr e, por último, Exc e Esf.

**Tabela 4.19:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[16,99 ; 17,01]	[19,17 ; 19,22]	[72,74 ; 74,07]	[16,17 ; 16,22]	[60,52 ; 60,71]	[16,05 ; 16,09]	[16,00 ; 16,05]
CP2	[32,01 ; 32,06]	[36,19 ; 36,26]	[86,29 ; 87,29]	[30,85 ; 30,91]	[68,46 ; 68,61]	[31,39 ; 31,46]	[31,25 ; 31,32]
CP3	[46,46 ; 46,52]	[52,58 ; 52,64]	[92,86 ; 93,51]	[44,85 ; 44,92]	[75,95 ; 76,07]	[46,22 ; 46,29]	[46,00 ; 46,07]
CP4	[60,50 ; 60,57]	[68,41 ; 68,47]	[96,36 ; 96,75]	[58,39 ; 58,46]	[83,13 ; 83,22]	[60,55 ; 60,62]	[60,26 ; 60,33]
CP5	[74,14 ; 74,19]	[83,76 ; 83,82]	[98,35 ; 98,55]	[71,42 ; 71,49]	[89,94 ; 90,01]	[74,36 ; 74,43]	[74,05 ; 74,12]
CP6	[87,26 ; 87,32]	[98,16 ; 98,21]	[99,50 ; 99,58]	[83,98 ; 84,03]	[96,32 ; 96,37]	[87,60 ; 87,66]	[87,29 ; 87,35]
CP7	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[95,97 ; 95,99]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP8	[100,00 ; 100,00]			[100,00 ; 100,00]			[100,00 ; 100,00]

**Tabela 4.20:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,0438	0,0674	0,0314	0,0691	0,0425	0,0575
Exc	0,0438	0,0000	0,0630	0,0532	0,0674	0,0653	0,0555
Ref	0,0674	0,0630	0,0000	0,0669	0,0497	0,0624	0,0583
Log	0,0314	0,0532	0,0669	0,0000	0,0707	0,0372	0,0569
Alr	0,0691	0,0674	0,0497	0,0707	0,0000	0,0537	0,0602
Clr	0,0425	0,0653	0,0624	0,0372	0,0537	0,0000	0,0541
Esf	0,0575	0,0555	0,0583	0,0569	0,0602	0,0541	0,0000



**Figura 4.10:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

#### 4.8.2 Uniforme

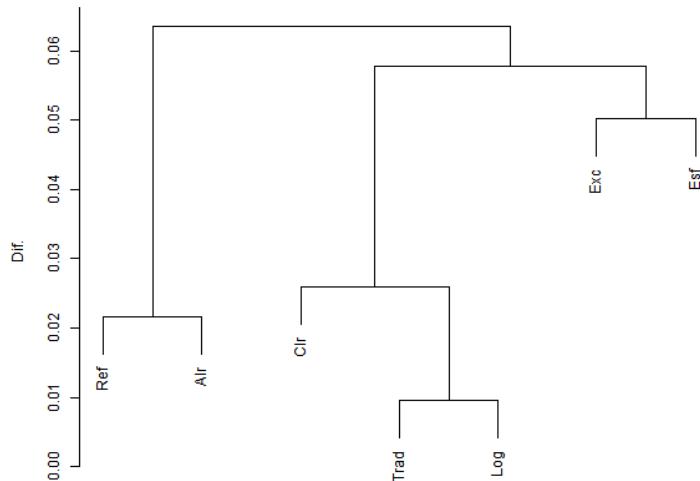
Para simular a estrutura da base composicional utilizando uma distribuição uniforme utilizamos dos seguintes vetores de médias e variâncias  $\mathbf{m} = (0, 3; 0, 1; 0, 1; 0, 1; 0, 1; 0, 1; 0, 1)$  e  $\mathbf{s}^2 = (0, 008; 0, 008; 0, 008; 0, 008; 0, 008; 0, 008; 0, 008)$ , respectivamente. Pela Tabela 4.21 notamos que as transformações Ref e Alr possuem variâncias explicadas muito similares, e também são as maiores médias, seguidas das transformações Clr, Exc, Trad, Log e Esf, respectivamente. Na Tabela 4.22 e Figura 4.11 notamos que as transformações Trad e Log possuem as interpretações das componentes principais similares, com a transformação Clr próxima também. As transformações Ref e Alr também levam a interpretações similares, seguidas da Exc e Esf.

**Tabela 4.21:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Clr	Esf
CP1	[21,93 ; 21,99]	[24,40 ; 24,48]	[59,05 ; 59,15]	[20,44 ; 20,51]	[56,71 ; 56,81]	[28,31 ; 28,47]	[18,32 ; 18,38]
CP2	[36,28 ; 36,35]	[40,61 ; 40,68]	[70,84 ; 70,93]	[34,78 ; 34,85]	[69,51 ; 69,60]	[41,89 ; 42,04]	[33,77 ; 33,84]
CP3	[50,05 ; 50,11]	[56,06 ; 56,13]	[78,34 ; 78,42]	[48,47 ; 48,54]	[77,19 ; 77,27]	[54,86 ; 54,98]	[48,54 ; 48,62]
CP4	[63,27 ; 63,34]	[70,92 ; 70,99]	[85,34 ; 85,41]	[61,70 ; 61,77]	[84,48 ; 84,54]	[67,31 ; 67,41]	[62,79 ; 62,86]
CP5	[76,03 ; 76,09]	[85,17 ; 85,23]	[91,96 ; 92,00]	[74,41 ; 74,48]	[91,34 ; 91,40]	[79,26 ; 79,35]	[76,52 ; 76,59]
CP6	[88,24 ; 88,30]	[99,00 ; 99,00]	[98,00 ; 98,00]	[86,70 ; 86,75]	[98,00 ; 98,00]	[90,63 ; 90,70]	[89,63 ; 89,69]
CP7	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[98,15 ; 98,19]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP8	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.22:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Clr	Esf
Trad	0,0000	0,0422	0,0594	0,0096	0,0588	0,0260	0,0576
Exc	0,0422	0,0000	0,0612	0,0431	0,0636	0,0528	0,0502
Ref	0,0594	0,0612	0,0000	0,0595	0,0216	0,0485	0,0561
Log	0,0096	0,0431	0,0595	0,0000	0,0593	0,0255	0,0577
Alr	0,0588	0,0636	0,0216	0,0593	0,0000	0,0452	0,0555
Clr	0,0260	0,0528	0,0485	0,0255	0,0452	0,0000	0,0489
Esf	0,0576	0,0502	0,0561	0,0577	0,0555	0,0489	0,0000

**Figura 4.11:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

### 4.8.3 Normal

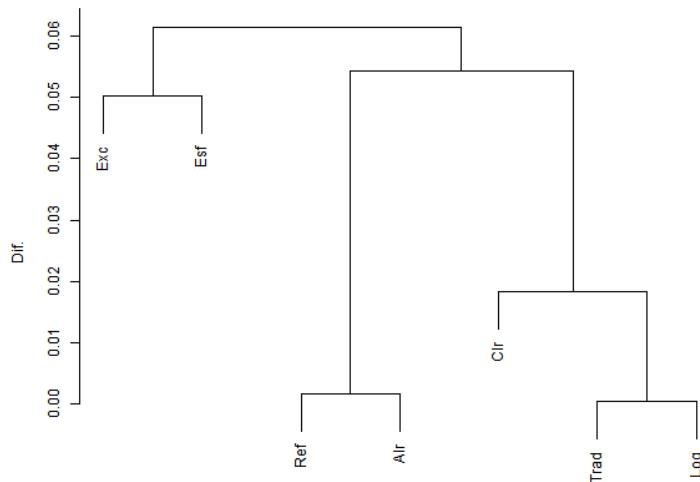
Assim como para a distribuição uniforme utilizamos  $\mathbf{m} = (0, 3; 0, 1; 0, 1; 0, 1; 0, 1; 0, 1; 0, 1)$  e  $\mathbf{s}^2 = (0, 008; 0, 008; 0, 008; 0, 008; 0, 008; 0, 008; 0, 008)$ . Pela Tabela 4.23 notamos que as transformações Ref e Alr possuem variâncias explicadas muito similares, e também são as maiores médias, seguidas das transformações Esf, Clr, Trad, Log e Exc, respectivamente. Na Tabela 4.24 e Figura 4.12 notamos que as transformações Trad e Log possuem as interpretações das componentes principais parecidas, com a transformação Clr próxima também. As transformações Ref e Alr também possuem interpretações similares das componentes principais, seguidas da Exc e Esf.

**Tabela 4.23:** Intervalo de Confiança para a média da proporção da variância acumulada explicada

	Trad	Exc	Ref	Log	Alr	Cir	Esf
CP1	[25,07 ; 25,14]	[23,36 ; 24,04]	[67,71 ; 69,82]	[25,07 ; 25,14]	[67,69 ; 69,80]	[27,00 ; 27,08]	[31,12 ; 32,50]
CP2	[38,88 ; 38,96]	[38,82 ; 39,52]	[77,10 ; 78,54]	[38,88 ; 38,96]	[77,09 ; 78,53]	[40,61 ; 40,69]	[45,42 ; 46,64]
CP3	[52,09 ; 52,16]	[53,64 ; 54,35]	[82,97 ; 84,06]	[52,09 ; 52,16]	[82,97 ; 84,06]	[53,59 ; 53,67]	[58,81 ; 59,86]
CP4	[64,78 ; 64,85]	[67,90 ; 68,60]	[88,39 ; 89,12]	[64,78 ; 64,85]	[88,39 ; 89,12]	[66,05 ; 66,12]	[71,40 ; 72,25]
CP5	[77,03 ; 77,09]	[81,62 ; 82,32]	[93,54 ; 93,93]	[77,03 ; 77,09]	[93,53 ; 93,92]	[78,02 ; 78,08]	[83,04 ; 83,65]
CP6	[88,85 ; 88,90]	[94,70 ; 95,39]	[98,62 ; 98,73]	[88,85 ; 88,90]	[98,62 ; 98,73]	[89,41 ; 89,48]	[93,35 ; 93,65]
CP7	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]	[100,00 ; 100,00]
CP8	[100,00 ; 100,00]			[100,00 ; 100,00]		[100,00 ; 100,00]	

**Tabela 4.24:** Matriz de distâncias entre transformações

	Trad	Exc	Ref	Log	Alr	Cir	Esf
Trad	0,0000	0,0444	0,0544	0,0005	0,0544	0,0183	0,0560
Exc	0,0444	0,0000	0,0615	0,0443	0,0615	0,0522	0,0504
Ref	0,0544	0,0615	0,0000	0,0544	0,0018	0,0440	0,0564
Log	0,0005	0,0443	0,0544	0,0000	0,0544	0,0182	0,0561
Alr	0,0544	0,0615	0,0018	0,0544	0,0000	0,0441	0,0563
Clr	0,0183	0,0522	0,0440	0,0182	0,0441	0,0000	0,0551
Esf	0,0560	0,0504	0,0564	0,0561	0,0563	0,0551	0,0000

**Figura 4.12:** Dendrograma da média da diferença absoluta entre as cargas das variáveis entre todas as componentes principais

## 4.9 Resumo das Simulações

As simulações realizadas geraram vários dendrogramas e tabelas, mas para melhor visualizarmos os resultados podemos resumí-los em dois tipos:

- Os dendrogramas que fazem parte do primeiro tipo formaram 3 grupos de transformações. São eles:

1. Grupo das transformações Trad, Log e Clr.
2. Grupo das transformações Ref e Alr.
3. Grupo das transformações Esf e Exc.

As simulações que geraram dendrogramas dessa forma são todas as simulações que possuíam uma variável com maior concentração tanto de 4 quanto de 8 variáveis (Figuras 4.4, 4.5, 4.6, 4.10, 4.11 e 4.12), como também as que foram simuladas com a mesma concentração em todas as variáveis pela distribuição Dirichlet, de 4 e 8 variáveis (Figuras 4.1 e 4.7), e pela Normal de 4 variáveis (Figura 4.3).

- Os dendrogramas que fazem parte do segundo tipo formaram 2 grupos de transformações. São eles:

1. Grupo das transformações Trad, Log e Clr
2. Grupo das transformações Ref, Alr

As transformações Exc e Esf não se encaixaram em nenhum grupo.

As simulações que geraram dendrogramas dessa forma são as que foram simuladas com a mesma concentração em todas as variáveis pela distribuição Uniforme, de 4 e 8 variáveis (Figuras 4.2 e 4.8), e pela Normal de 8 variáveis (Figura 4.9).

As transformações Exc e Esf foram as mais diferentes, pois mesmo quando formaram um grupo, este foi o último a ser formado. No caso do segundo tipo de dendrograma, se elas fossem agrupadas às outras transformações, Exc ficaria junto com Trad, Log e Clr e Esf com Ref e Alr, mas em patamares mais altos na medida de dissimilaridade.

# Capítulo 5

## Conclusões

Com as 7 transformações aplicadas nas mesmas bases de dados pudemos comparar o desempenho de cada uma, dadas algumas condições de número de variáveis e também de número de observações de cada amostra.

Notamos que na base de dados 1 praticamente todas as transformações apresentaram resultados similares, tanto com relação à proporção da variância explicada quanto às cargas das variáveis, mas pudemos notar pelas figuras das coordenadas das variáveis que as transformações Trad, Exc e Log, as transformações Ref e Alr e as transformações Clr e Esf são mais parecidas entre si.

Na base 2 notamos que todas as transformações tiveram praticamente o mesmo resultado. Isso nos indica que quando uma base de dados possuir poucas variáveis, independentemente da transformação, os resultados serão similares, principalmente as transformações Trad, Log e Clr resultam em coordenadas similares, assim como as transformações Ref e Alr e também Exc e Esf.

Na base 3 notamos que, com muitas variáveis, os resultados das transformações são muito diversos, podendo alterar significativamente a interpretação dos resultados, sendo necessário um maior cuidado na hora da escolha da transformação utilizada. Entretanto pudemos notar um comportamento semelhante entre as cargas das variáveis quando foram aplicados as transformações Ref e Alr, assim como nas transformações Trad, Log e Exc.

Com base nas análises de simulações realizadas concluímos primeiramente que a estrutura de simulação dos dados (Dirichlet, Uniforme e Normal) não afeta os resultados obtidos. Com isso vemos que as transformações Ref e Alr têm a interpretação das componentes principais similares assim como o desempenho na explicação da proporção da variância. Pelos mesmos critérios notamos que as transformações Log e Trad são similares, sendo que a transformação Clr também está próxima dessas transformações, principalmente no critério da interpretação das componentes principais. Já as transformações Exc e Esf são voláteis tanto no desempenho da explicação da variância como na interpretação das componentes

principais.

Como podemos notar tanto na análise dos 3 bancos de dados reais como nas análises de simulações as transformações Ref e Alr e as transformações Trad e Log são parecidas o que indica que a aplicação do logaritmo (Alr e Log) praticamente não altera a interpretação das componentes principais e a explicação da variância. Também concluímos que as transformações em que é escolhida uma variável de referência (Ref e Alr) são as com melhor desempenho na explicação da variância, porque têm uma variância a menos a ser explicada.

## Apêndice A

# Resultados da Análise de Componentes Principais

Este apêndice contém todos os resultados de cada análise de componentes principais realizada.

## A.1 Análise base 1

### A.1.1 Análise Tradicional (Trad)

**Tabela A.1:** Variância das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Variância	3,40	1,60	1,00	0,53	0,38	0,09	0,00
Proporção da Variância	0,49	0,23	0,14	0,08	0,05	0,01	0,00
Proporção Acumulada da Variância	0,49	0,71	0,86	0,93	0,99	1,00	1,00

**Tabela A.2:** Cargas das Componentes principais

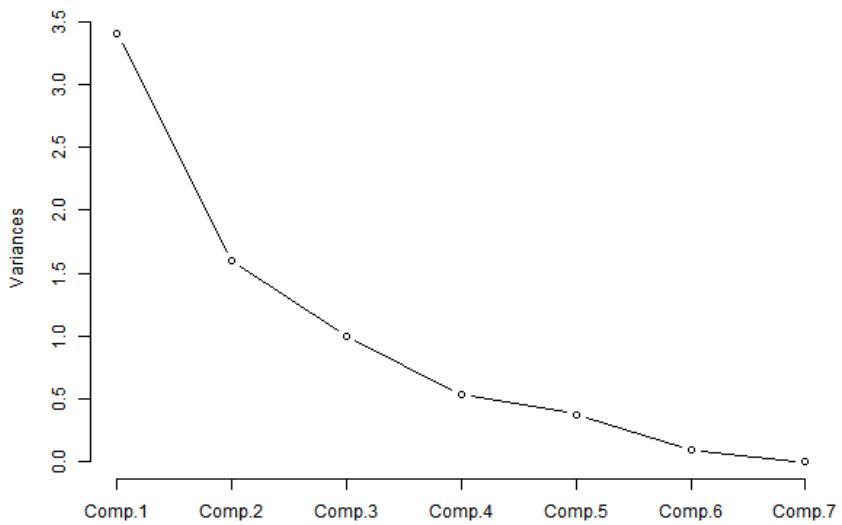
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	0,52	-0,09	0,05	0,28	-0,08	0,50	0,62
Si	-0,53	-0,13	0,00	-0,09	0,09	-0,34	0,75
Al	-0,07	0,70	0,13	-0,21	-0,65	0,03	0,14
Mg	0,40	0,25	0,00	-0,72	0,49	-0,07	0,15
Mn	0,49	0,13	-0,20	0,34	-0,09	-0,76	0,08
K	-0,10	0,14	-0,96	-0,03	-0,00	0,20	0,04
Ti	-0,19	0,62	0,12	0,49	0,56	0,10	0,01

**Tabela A.3:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

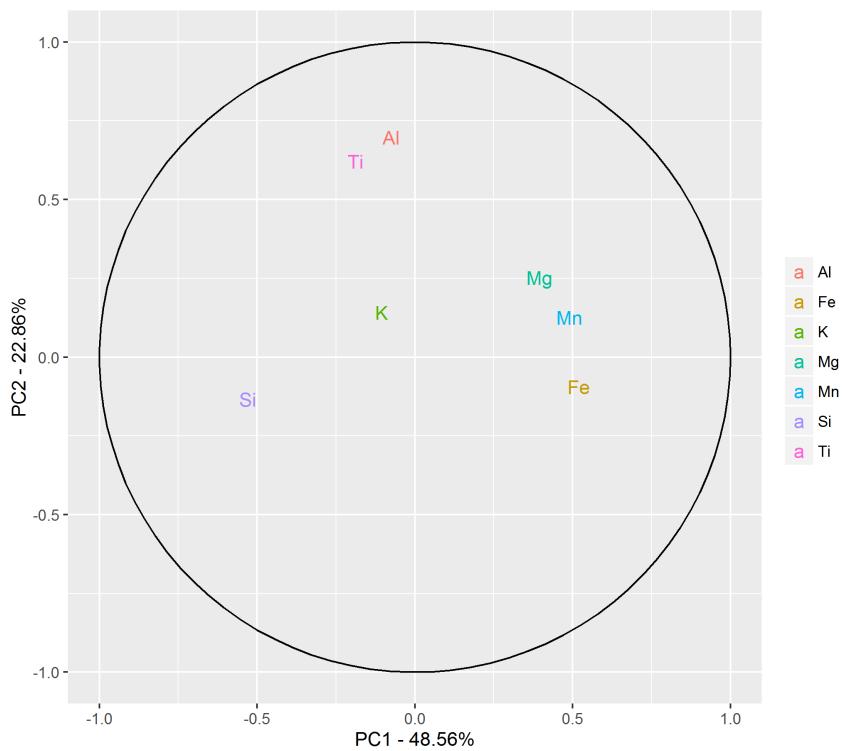
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	91,74	93,14	93,41	97,46	97,69	99,99	100,00
Si	95,35	98,20	98,20	98,62	98,95	100,00	100,00
Al	1,86	79,82	81,59	83,99	99,98	99,99	100,00
Mg	53,29	63,50	63,50	90,92	99,95	100,00	100,00
Mn	81,91	84,43	88,25	94,47	94,81	100,00	100,00
K	3,73	6,96	99,61	99,65	99,65	100,00	100,00
Ti	12,07	73,88	75,24	88,03	99,90	100,00	100,00

**Tabela A.4:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	0,96	-0,12	0,05	0,20	-0,05	0,15	0,00
Si	-0,98	-0,17	0,00	-0,06	0,06	-0,10	0,00
Al	-0,14	0,88	0,13	-0,16	-0,40	0,01	0,00
Mg	0,73	0,32	0,00	-0,52	0,30	-0,02	0,00
Mn	0,91	0,16	-0,20	0,25	-0,06	-0,23	0,00
K	-0,19	0,18	-0,96	-0,02	-0,00	0,06	0,00
Ti	-0,35	0,79	0,12	0,36	0,34	0,03	0,00



**Figura A.1:** Proporção da variância explicada



**Figura A.2:** Coordenadas das variáveis

### A.1.2 Exclusão de uma Variável (Exc)

**Tabela A.5:** *Variâncias das Componentes principais*

	CP1	CP2	CP3	CP4	CP5	CP6
Variância	3,32	1,16	0,97	0,46	0,10	0,00
Proporção da Variância	0,55	0,19	0,16	0,08	0,02	0,00
Proporção Acumulada da Variância	0,55	0,75	0,91	0,98	1,00	1,00

**Tabela A.6:** *Cargas das Componentes principais*

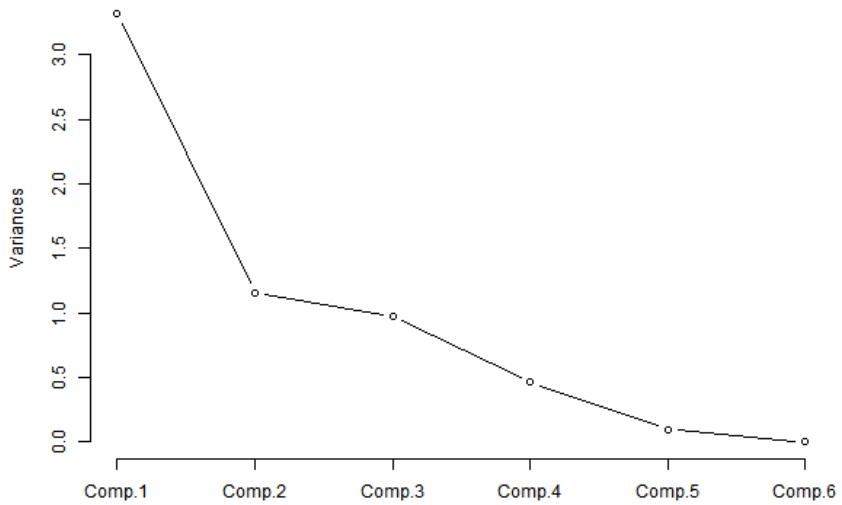
	CP1	CP2	CP3	CP4	CP5	CP6
Fe	0,52	-0,22	0,05	0,23	0,50	0,61
Si	-0,54	-0,06	0,01	-0,12	-0,34	0,76
Al	-0,03	0,79	-0,48	0,34	0,10	0,15
Mg	0,41	0,28	-0,13	-0,84	-0,08	0,15
Mn	0,50	0,08	0,21	0,33	-0,76	0,08
K	-0,10	0,49	0,84	-0,04	0,20	0,04

**Tabela A.7:** *Proporção da variância acumulada de cada variável explicada pelas componentes principais*

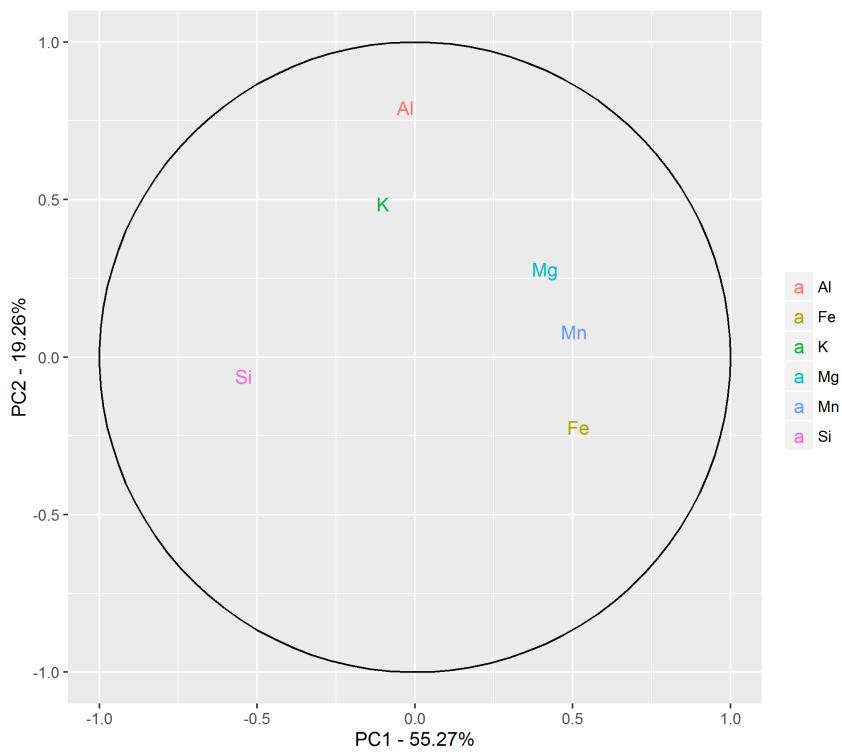
	CP1	CP2	CP3	CP4	CP5	CP6
Fe	89,13	94,92	95,18	97,59	99,99	100,00
Si	97,75	98,18	98,19	98,91	100,00	100,00
Al	0,29	72,64	94,56	99,90	99,99	100,00
Mg	56,57	65,58	67,31	99,93	100,00	100,00
Mn	84,50	85,23	89,36	94,44	100,00	100,00
K	3,39	30,65	99,55	99,61	100,00	100,00

**Tabela A.8:** *Matriz de correlação entre as variáveis e as componentes principais*

	CP1	CP2	CP3	CP4	CP5	CP6
Fe	0,94	-0,24	0,05	0,16	0,15	0,01
Si	-0,99	-0,07	0,01	-0,08	-0,10	0,01
Al	-0,05	0,85	-0,47	0,23	0,03	0,00
Mg	0,75	0,30	-0,13	-0,57	-0,03	0,00
Mn	0,92	0,09	0,20	0,23	-0,24	0,00
K	-0,18	0,52	0,83	-0,02	0,06	0,00



**Figura A.3:** Proporção da variância explicada



**Figura A.4:** Coordenadas das variáveis

### A.1.3 Variável de Referência (Ref)

**Tabela A.9:** *Variâncias das Componentes principais*

	CP1	CP2	CP3	CP4	CP5	CP6
Variância	3,26	1,35	0,70	0,44	0,20	0,06
Proporção da Variância	0,54	0,23	0,12	0,07	0,03	0,01
Proporção Acumulada da Variância	0,54	0,77	0,89	0,96	0,99	1,00

**Tabela A.10:** *Cargas das Componentes principais*

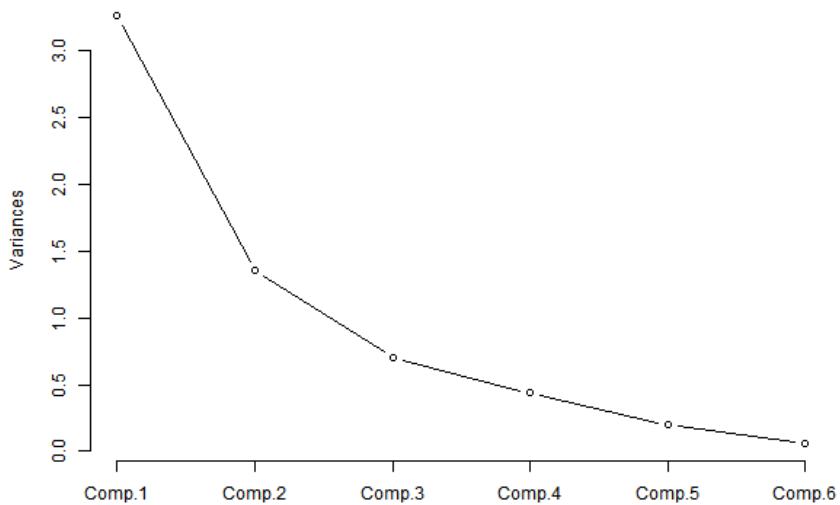
	CP1	CP2	CP3	CP4	CP5	CP6
Fe	0,51	0,12	-0,09	0,46	0,27	0,66
Si	-0,16	-0,72	-0,33	0,53	0,12	-0,21
Al	0,39	-0,31	-0,48	-0,66	0,28	0,03
Mg	0,51	-0,07	-0,16	0,09	-0,83	-0,12
Mn	0,51	0,20	0,20	0,19	0,37	-0,70
K	0,21	-0,56	0,77	-0,16	-0,01	0,16

**Tabela A.11:** *Proporção da variância acumulada de cada variável explicada pelas componentes principais*

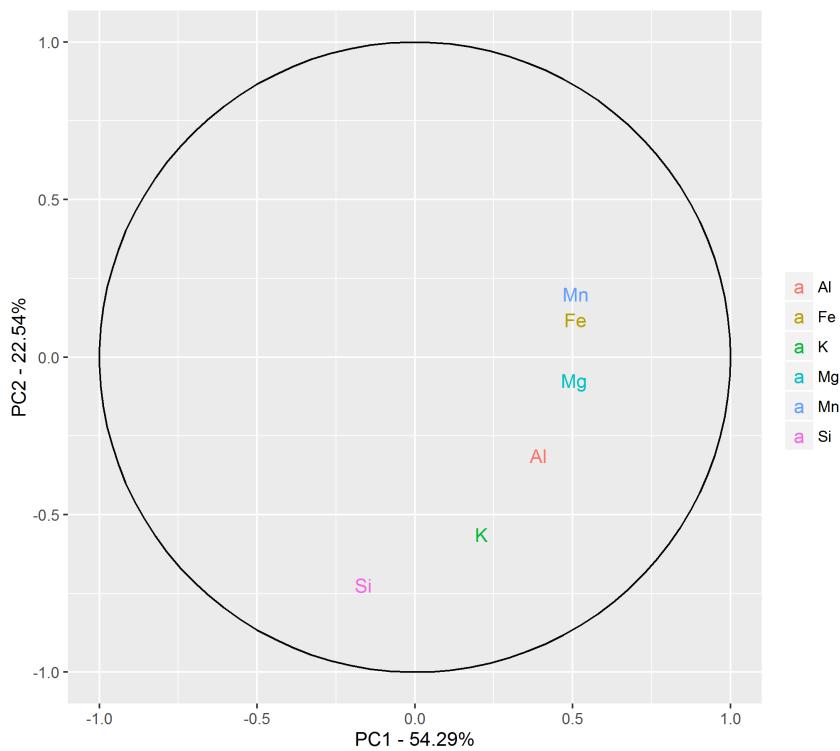
	CP1	CP2	CP3	CP4	CP5	CP6
Fe	84,45	86,32	86,88	96,11	97,59	100,00
Si	8,61	79,53	87,25	99,46	99,74	100,00
Al	49,97	63,23	79,41	98,39	99,98	100,00
Mg	83,35	84,11	86,00	86,36	99,92	100,00
Mn	84,79	90,21	92,93	94,55	97,30	100,00
K	14,58	57,57	98,73	99,86	99,86	100,00

**Tabela A.12:** *Matriz de correlação entre as variáveis e as componentes principais*

	CP1	CP2	CP3	CP4	CP5	CP6
Fe	0,92	0,14	-0,07	0,30	0,12	0,16
Si	-0,29	-0,84	-0,28	0,35	0,05	-0,05
Al	0,71	-0,36	-0,40	-0,44	0,13	0,01
Mg	0,91	-0,09	-0,14	0,06	-0,37	-0,03
Mn	0,92	0,23	0,16	0,13	0,17	-0,16
K	0,38	-0,66	0,64	-0,11	-0,01	0,04



**Figura A.5:** Proporção da variância explicada



**Figura A.6:** Coordenadas das variáveis

### A.1.4 Transformação Log (Log)

**Tabela A.13:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Variância	3,43	1,61	0,96	0,51	0,37	0,10	0,02
Proporção da Variância	0,49	0,23	0,14	0,07	0,05	0,01	0,00
Proporção Acumulada da Variância	0,49	0,72	0,86	0,93	0,98	1,00	1,00

**Tabela A.14:** Cargas das Componentes principais

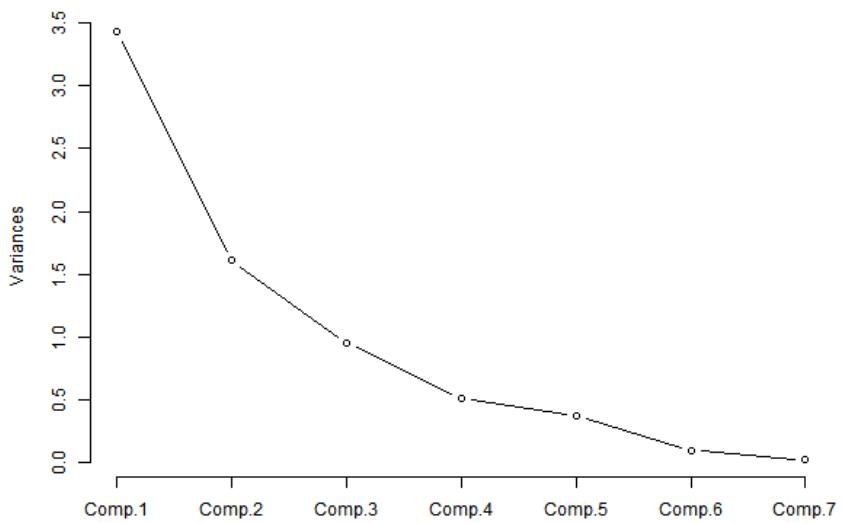
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	0,52	-0,06	0,03	0,33	-0,10	0,13	0,77
Si	-0,51	-0,17	-0,01	-0,06	-0,06	-0,70	0,46
Al	-0,08	0,69	0,20	-0,17	-0,67	0,02	0,08
Mg	0,40	0,26	0,04	-0,72	0,42	-0,23	0,15
Mn	0,50	0,10	-0,15	0,34	-0,16	-0,66	-0,38
K	-0,10	0,25	-0,95	-0,02	0,02	0,10	0,11
Ti	-0,21	0,60	0,18	0,47	0,58	-0,04	0,06

**Tabela A.15:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

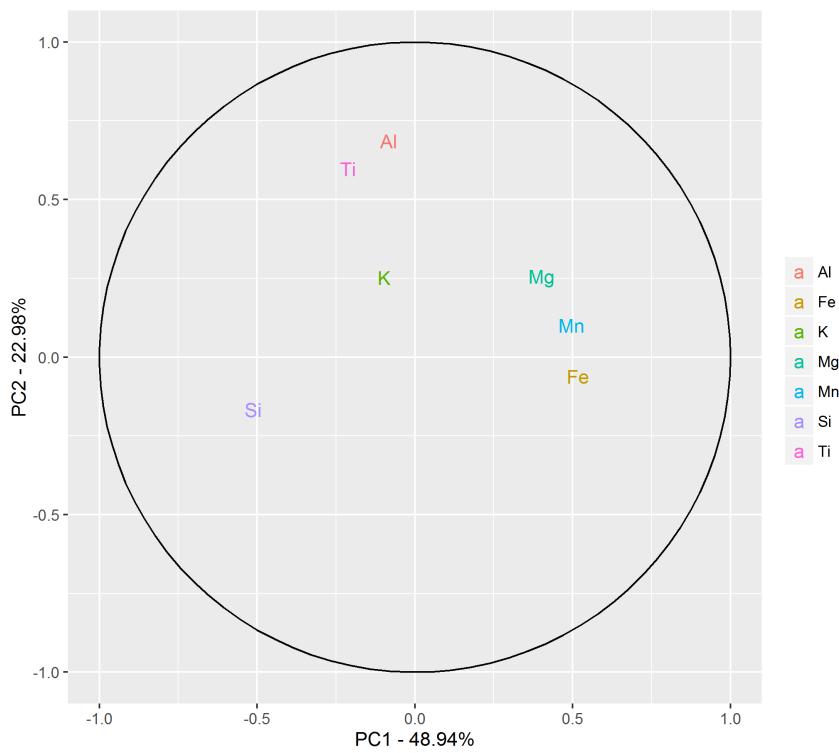
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	91,65	92,24	92,31	98,01	98,40	98,56	100,00
Si	89,88	94,36	94,37	94,55	94,70	99,47	100,00
Al	2,34	78,05	81,73	83,14	99,97	99,98	100,00
Mg	55,47	66,06	66,24	92,95	99,40	99,92	100,00
Mn	84,71	86,35	88,45	94,48	95,40	99,64	100,00
K	3,21	13,50	99,84	99,86	99,87	99,97	100,00
Ti	15,30	72,82	76,02	87,25	99,97	99,99	100,00

**Tabela A.16:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	0,96	-0,08	0,03	0,24	-0,06	0,04	0,12
Si	-0,95	-0,21	-0,01	-0,04	-0,04	-0,22	0,07
Al	-0,15	0,87	0,19	-0,12	-0,41	0,01	0,01
Mg	0,74	0,33	0,04	-0,52	0,25	-0,07	0,02
Mn	0,92	0,13	-0,15	0,25	-0,10	-0,21	-0,06
K	-0,18	0,32	-0,93	-0,01	0,01	0,03	0,02
Ti	-0,39	0,76	0,18	0,34	0,36	-0,01	0,01



**Figura A.7:** Proporção da variância explicada



**Figura A.8:** Coordenadas das variáveis

### A.1.5 *Additive Log-Ratio (Alr)*

**Tabela A.17:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6
Variância	3,19	1,39	0,74	0,41	0,22	0,04
Proporção da Variância	0,53	0,23	0,12	0,07	0,04	0,01
Proporção Acumulada da Variância	0,53	0,76	0,89	0,96	0,99	1,00

**Tabela A.18:** Cargas das Componentes principais

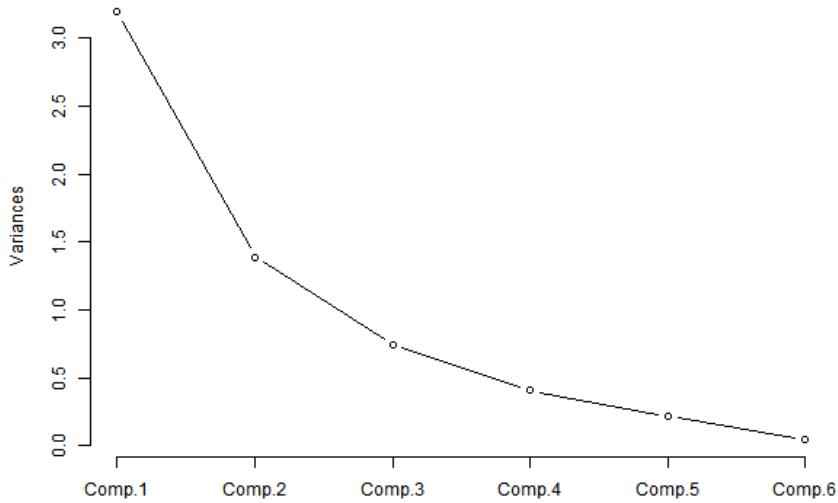
	CP1	CP2	CP3	CP4	CP5	CP6
Fe	0,52	0,11	0,07	0,41	-0,27	0,68
Si	-0,18	-0,70	0,31	0,57	-0,15	-0,18
Al	0,37	-0,39	0,44	-0,67	-0,25	0,04
Mg	0,50	-0,09	0,17	0,13	0,82	-0,15
Mn	0,52	0,20	-0,16	0,17	-0,41	-0,68
K	0,18	-0,54	-0,80	-0,12	0,03	0,11

**Tabela A.19:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

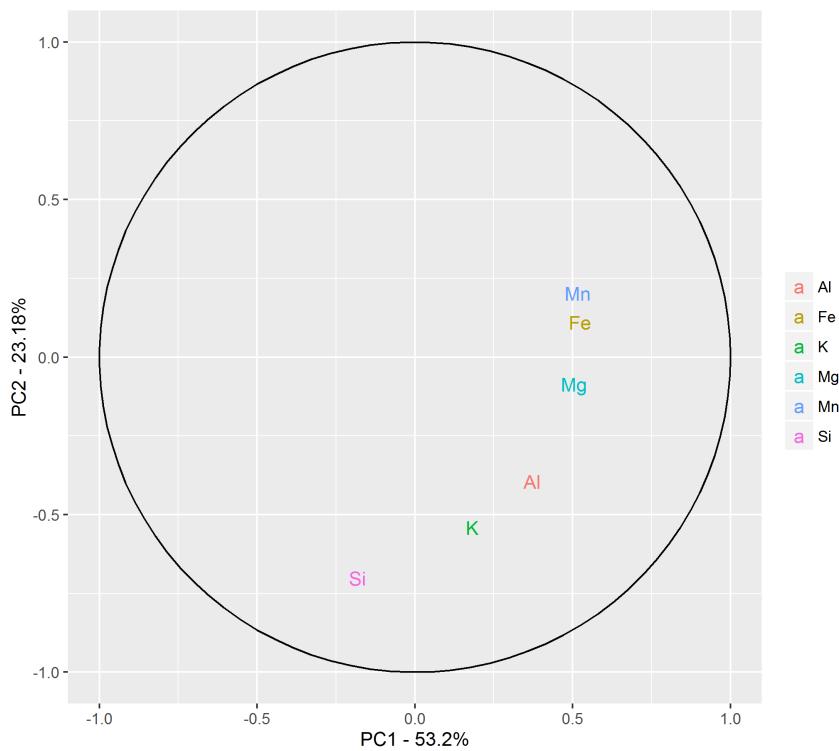
	CP1	CP2	CP3	CP4	CP5	CP6
Fe	87,36	89,04	89,39	96,31	97,92	100,00
Si	10,50	78,89	86,25	99,34	99,85	100,00
Al	44,10	65,75	80,38	98,67	99,99	100,00
Mg	81,27	82,32	84,42	85,15	99,90	100,00
Mn	85,36	91,04	92,90	94,14	97,92	100,00
K	10,63	51,28	99,34	99,92	99,95	100,00

**Tabela A.20:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	0,96	-0,08	0,03	0,24	-0,06	0,04	0,12
Si	-0,95	-0,21	-0,01	-0,04	-0,04	-0,22	0,07
Al	-0,15	0,87	0,19	-0,12	-0,41	0,01	0,01
Mg	0,74	0,33	0,04	-0,52	0,25	-0,07	0,02
Mn	0,92	0,13	-0,15	0,25	-0,10	-0,21	-0,06
K	-0,18	0,32	-0,93	-0,01	0,01	0,03	0,02
Ti	-0,39	0,76	0,18	0,34	0,36	-0,01	0,01



**Figura A.9:** Proporção da variância explicada



**Figura A.10:** Coordenadas das variáveis

### A.1.6 *Center Log-Ratio* (Clr)

**Tabela A.21:** *Variâncias das Componentes principais*

	CP1	CP2	CP3	CP4	CP5	CP6
Variância	3,73	1,52	0,70	0,56	0,39	0,10
Proporção da Variância	0,53	0,22	0,10	0,08	0,06	0,01
Proporção Acumulada da Variância	0,53	0,75	0,85	0,93	0,99	1,00

**Tabela A.22:** *Cargas das Componentes principais*

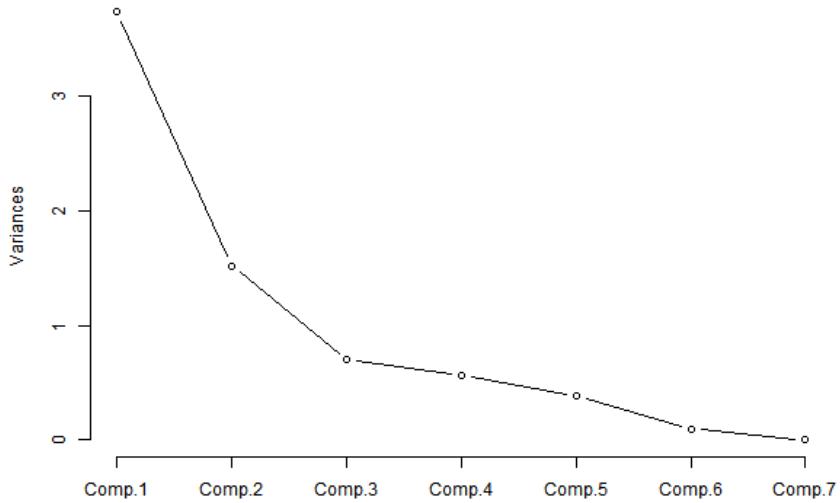
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	0,48	-0,13	-0,25	0,20	-0,04	0,74	-0,33
Si	-0,44	-0,00	0,00	0,70	-0,11	-0,15	-0,53
Al	-0,32	-0,48	-0,01	-0,42	-0,66	0,12	-0,19
Mg	0,25	-0,45	0,80	0,02	0,24	-0,05	-0,20
Mn	0,49	0,06	-0,26	-0,24	-0,06	-0,57	-0,55
K	-0,20	0,66	0,35	-0,37	0,02	0,29	-0,41
Ti	-0,38	-0,33	-0,32	-0,30	0,70	0,08	-0,26

**Tabela A.23:** *Proporção da variância acumulada de cada variável explicada pelas componentes principais*

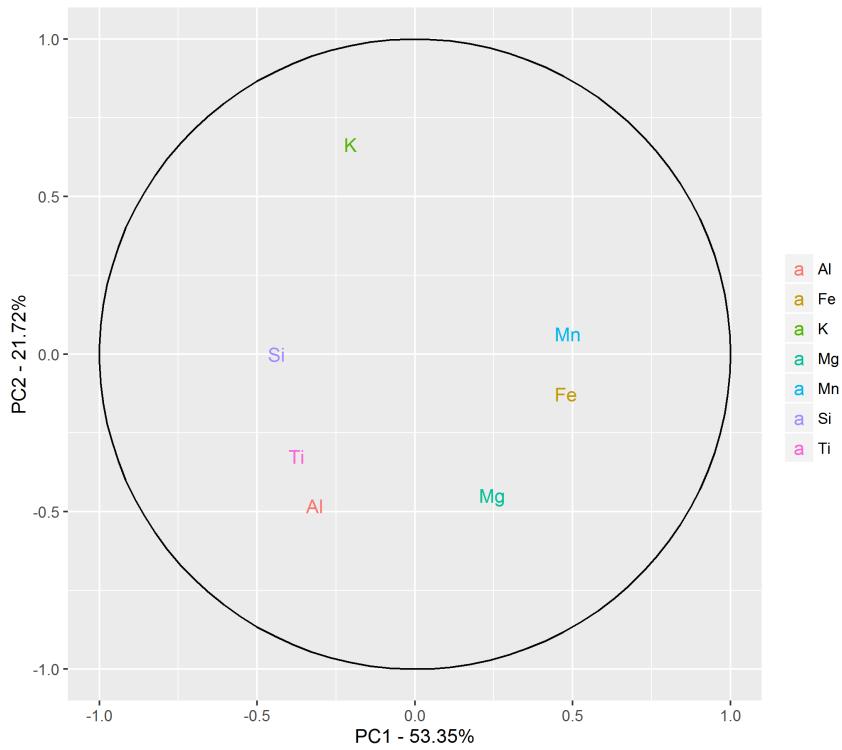
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	85,46	87,94	92,35	94,64	94,70	100,00	100,00
Si	71,63	71,63	71,63	99,30	99,78	100,00	100,00
Al	37,71	72,91	72,91	82,96	99,88	100,00	100,00
Mg	22,44	52,97	97,78	97,81	99,97	99,99	100,00
Mn	88,04	88,65	93,37	96,61	96,77	100,00	100,00
K	15,51	82,64	91,40	99,16	99,18	99,99	100,00
Ti	52,69	68,77	76,11	81,04	99,93	100,00	100,00

**Tabela A.24:** *Matriz de correlação entre as variáveis e as componentes principais*

	CP1	CP2	CP3	CP4	CP5	CP6
Fe	0,93	0,13	0,06	0,26	-0,13	0,14
Si	-0,32	-0,83	0,27	0,36	-0,07	-0,04
Al	0,66	-0,47	0,38	-0,43	-0,12	0,01
Mg	0,90	-0,10	0,14	0,09	0,38	-0,03
Mn	0,92	0,24	-0,14	0,11	-0,19	-0,14
K	0,33	-0,64	-0,69	-0,08	0,02	0,02



**Figura A.11:** Proporção da variância explicada



**Figura A.12:** Coordenadas das variáveis

### A.1.7 Transformação Esférica (Esf)

**Tabela A.25:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6
Variância	3,25	1,52	0,70	0,33	0,13	0,05
Proporção da Variância	0,54	0,25	0,12	0,06	0,02	0,01
Proporção Acumulada da Variância	0,54	0,80	0,91	0,97	0,99	1,00

**Tabela A.26:** Cargas das Componentes principais

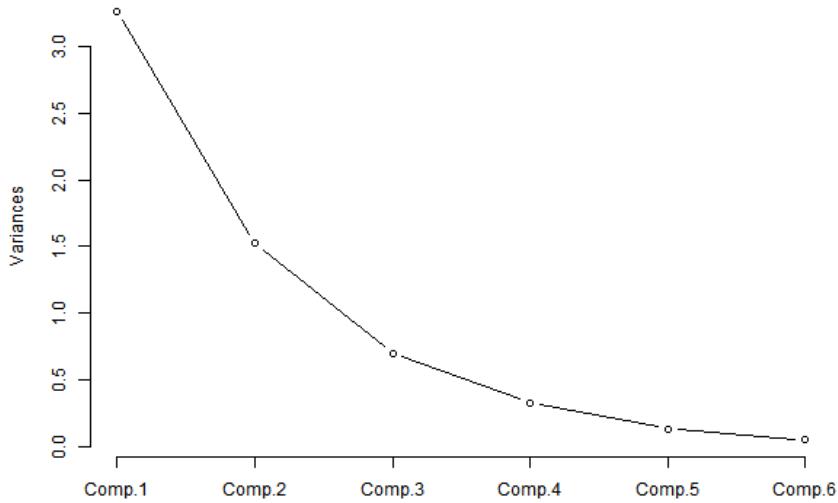
	CP1	CP2	CP3	CP4	CP5	CP6
Fe	0,54	-0,06	-0,11	-0,05	-0,29	0,78
Si	-0,49	0,06	0,03	-0,76	-0,40	0,15
Al	-0,44	-0,13	-0,58	0,50	-0,45	0,08
Mg	-0,13	-0,63	0,67	0,22	-0,30	0,04
Mn	0,51	-0,19	-0,21	-0,21	-0,51	-0,60
K	0,03	0,74	0,40	0,28	-0,46	-0,06

**Tabela A.27:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

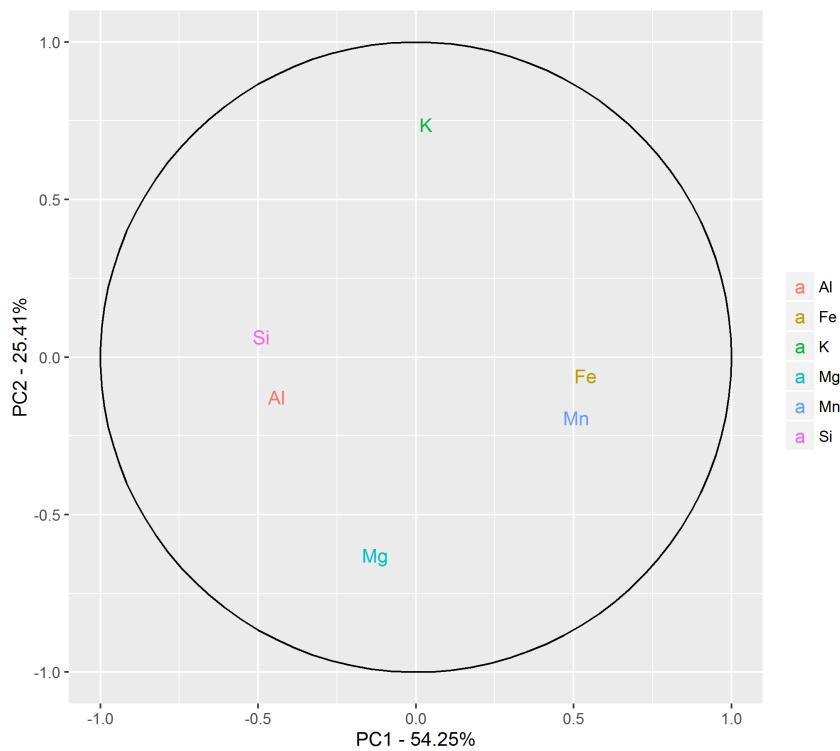
	CP1	CP2	CP3	CP4	CP5	CP6
Fe	94,20	94,74	95,59	95,67	96,82	100,00
Si	78,04	78,66	78,71	97,77	99,88	100,00
Al	63,12	65,61	88,95	97,28	99,97	100,00
Mg	5,41	65,81	97,18	98,76	100,00	100,00
Mn	84,37	90,05	93,27	94,67	98,11	100,00
K	0,34	83,07	94,53	97,15	99,98	100,00

**Tabela A.28:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
Fe	0,92	-0,16	-0,21	0,15	-0,03	0,23	-0,00
Si	-0,85	-0,00	0,00	0,53	-0,07	-0,05	-0,00
Al	-0,61	-0,59	-0,00	-0,32	-0,41	0,04	-0,00
Mg	0,47	-0,55	0,67	0,02	0,15	-0,01	-0,00
Mn	0,94	0,08	-0,22	-0,18	-0,04	-0,18	-0,00
K	-0,39	0,82	0,30	-0,28	0,01	0,09	-0,00
Ti	-0,73	-0,40	-0,27	-0,22	0,43	0,03	-0,00



**Figura A.13:** Proporção da variância explicada



**Figura A.14:** Coordenadas das variáveis

## A.2 Análise base 2

### A.2.1 Análise Tradicional (Trad)

**Tabela A.29:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4
Variância	3,40	0,34	0,26	0,00
Proporção da Variância	0,85	0,09	0,07	0,00
Proporção Acumulada da Variância	0,85	0,93	1,00	1,00

**Tabela A.30:** Cargas das Componentes principais

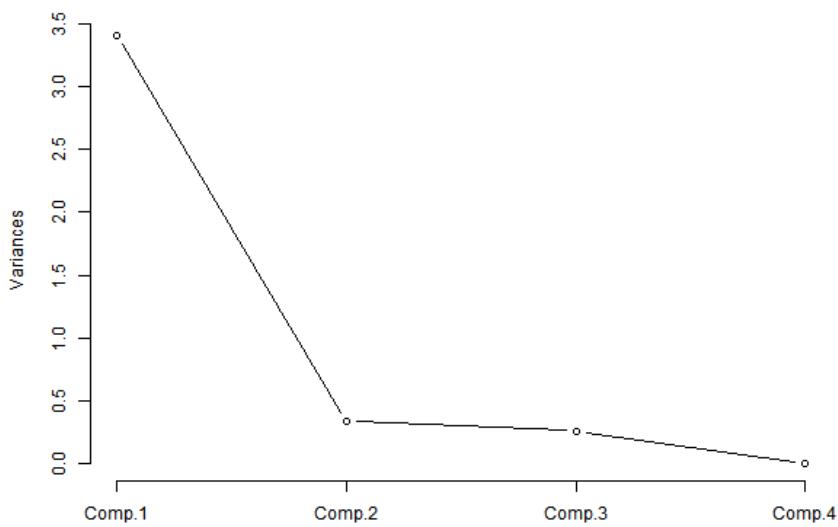
	CP1	CP2	CP3	CP4
Qm	0,52	0,52	-0,09	0,67
Rf	-0,51	-0,03	0,70	0,50
Qp	-0,50	-0,22	-0,70	0,46
M	0,47	-0,83	0,10	0,29

**Tabela A.31:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

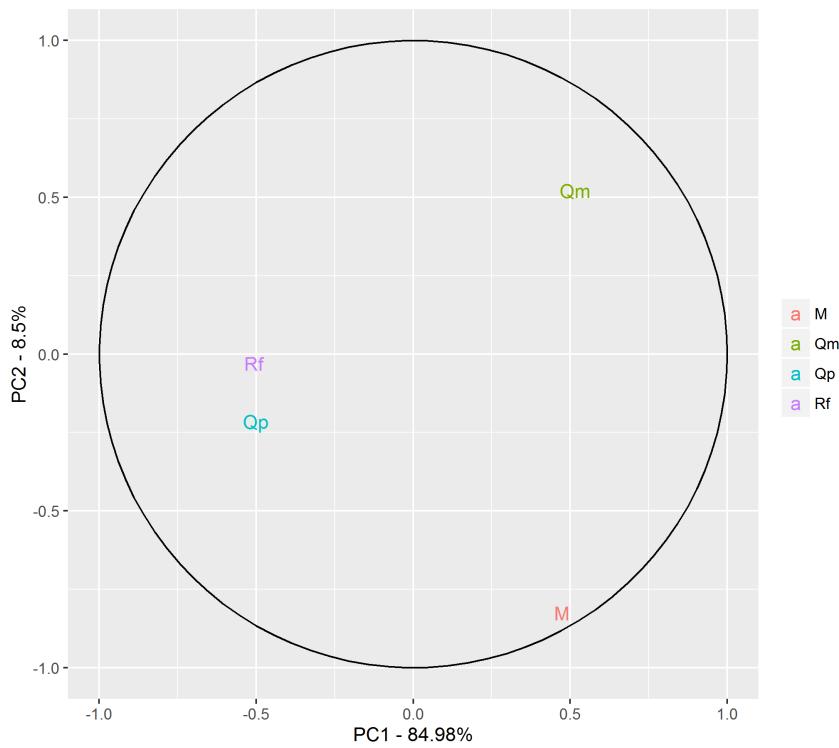
	CP1	CP2	CP3	CP4
Qm	90,58	99,80	100,00	100,00
Rf	87,19	87,22	100,00	100,00
Qp	85,63	87,21	100,00	100,00
M	76,54	99,73	100,00	100,00

**Tabela A.32:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4
Qm	0,95	0,30	-0,04	-0,00
Rf	-0,93	-0,02	0,36	-0,00
Qp	-0,93	-0,13	-0,36	-0,00
M	0,87	-0,48	0,05	-0,00



**Figura A.15:** Proporção da variância explicada



**Figura A.16:** Coordenadas das variáveis

### A.2.2 Exclusão de uma Variável (Exc)

**Tabela A.33:** *Variâncias das Componentes principais*

	CP1	CP2	CP3
Variância	2,70	0,26	0,04
Proporção da Variância	0,90	0,09	0,01
Proporção Acumulada da Variância	0,90	0,99	1,00

**Tabela A.34:** *Cargas das Componentes principais*

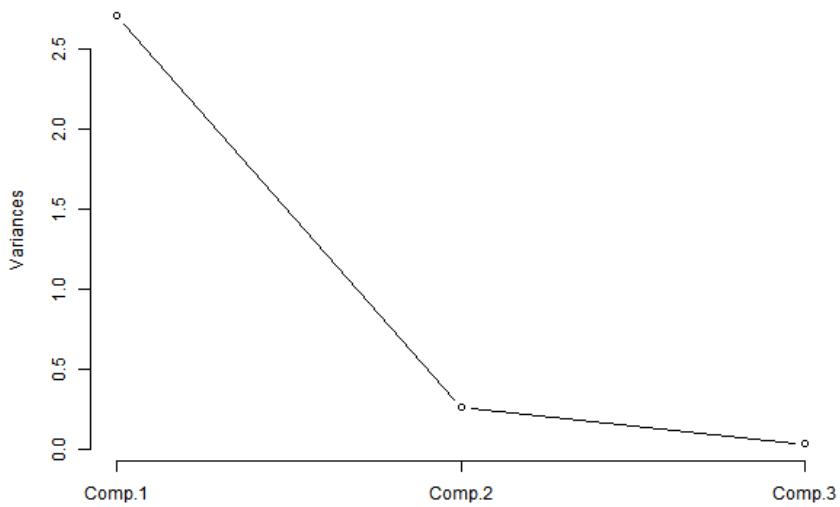
	CP1	CP2	CP3
Qm	0,60	0,01	0,80
Rf	-0,57	-0,70	0,44
Qp	-0,56	0,72	0,41

**Tabela A.35:** *Proporção da variância acumulada de cada variável explicada pelas componentes principais*

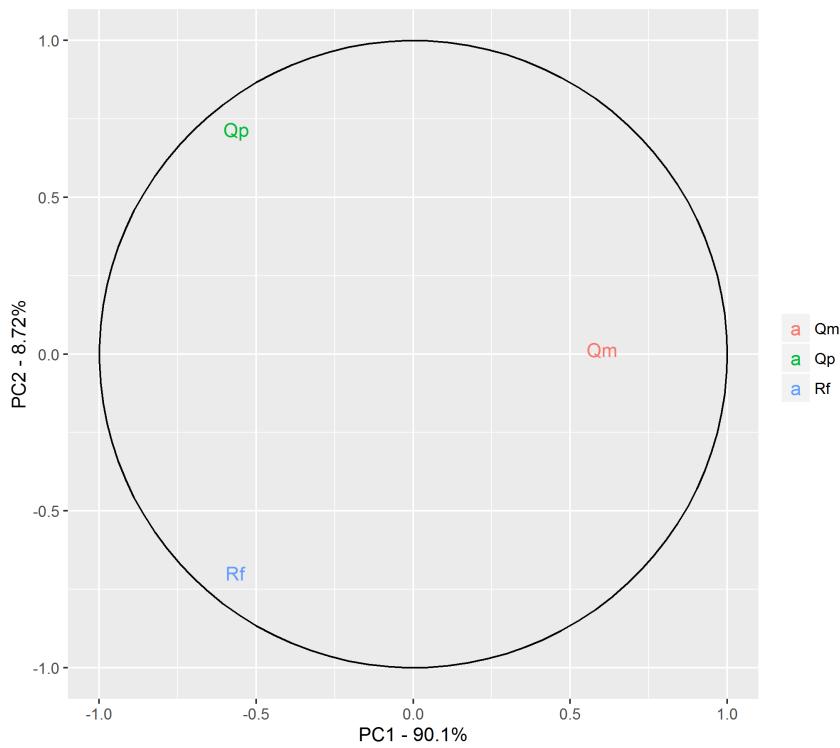
	CP1	CP2	CP3
Qm	97,75	97,75	100,00
Rf	86,55	99,33	100,00
Qp	86,01	99,40	100,00

**Tabela A.36:** *Matriz de correlação entre as variáveis e as componentes principais*

	CP1	CP2	CP3
Qm	0,99	0,01	0,15
Rf	-0,93	-0,36	0,08
Qp	-0,93	0,37	0,08



**Figura A.17:** Proporção da variância explicada



**Figura A.18:** Coordenadas das variáveis

### A.2.3 Variável de Referência (Ref)

**Tabela A.37:** *Variâncias das Componentes principais*

	CP1	CP2	CP3
Variância	2,88	0,10	0,02
Proporção da Variância	0,96	0,03	0,01
Proporção Acumulada da Variância	0,96	0,99	1,00

**Tabela A.38:** *Cargas das Componentes principais*

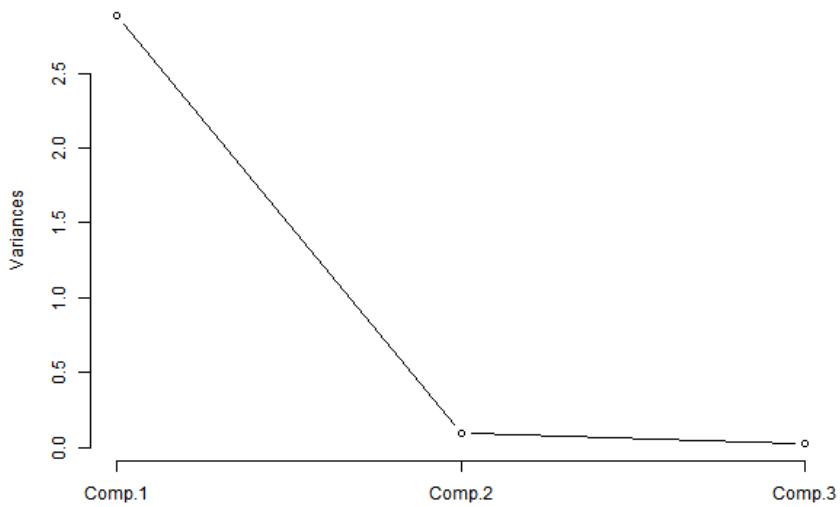
	CP1	CP2	CP3
Qm	-0,57	0,76	0,31
Rf	-0,58	-0,11	-0,80
Qp	-0,58	-0,64	0,51

**Tabela A.39:** *Proporção da variância acumulada de cada variável explicada pelas componentes principais*

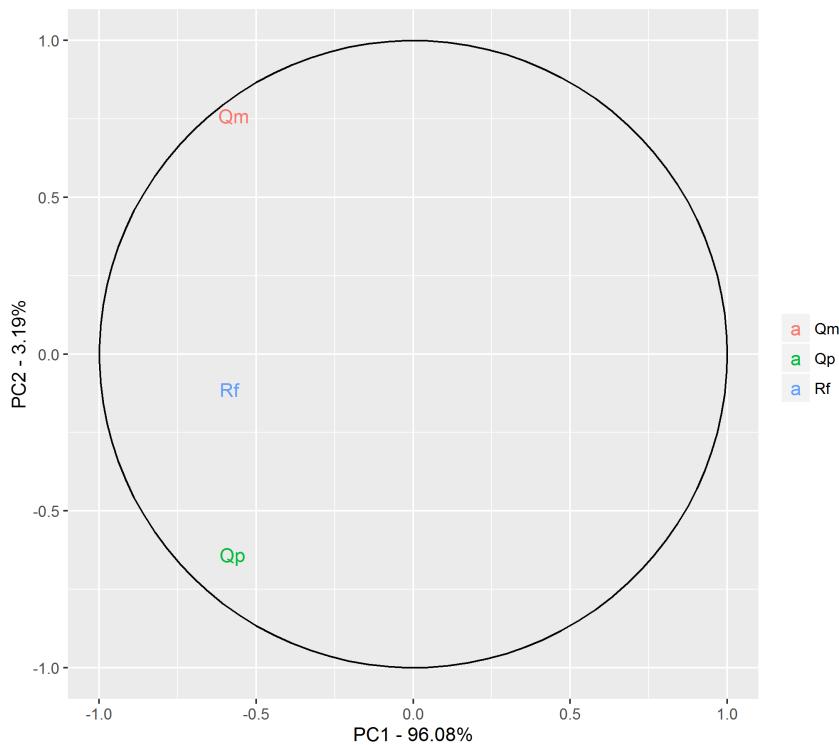
	CP1	CP2	CP3
Qm	94,27	99,79	100,00
Rf	98,46	98,58	100,00
Qp	95,51	99,43	100,00

**Tabela A.40:** *Matriz de correlação entre as variáveis e as componentes principais*

	CP1	CP2	CP3
Qm	-0,97	0,23	0,05
Rf	-0,99	-0,03	-0,12
Qp	-0,98	-0,20	0,08



**Figura A.19:** Proporção da variância explicada



**Figura A.20:** Coordenadas das variáveis

### A.2.4 Transformação Log (Log)

**Tabela A.41:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4
Variância	3,18	0,50	0,18	0,14
Proporção da Variância	0,79	0,12	0,05	0,04
Proporção Acumulada da Variância	0,79	0,92	0,96	1,00

**Tabela A.42:** Cargas das Componentes principais

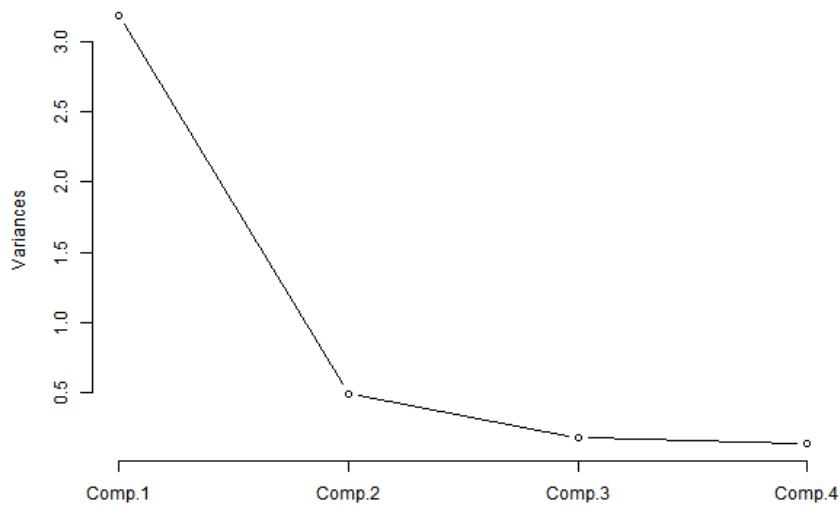
	CP1	CP2	CP3	CP4
Qm	0,51	0,40	0,31	0,69
Rf	-0,53	-0,12	-0,49	0,68
Qp	-0,52	-0,22	0,81	0,15
M	0,44	-0,88	0,00	0,18

**Tabela A.43:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

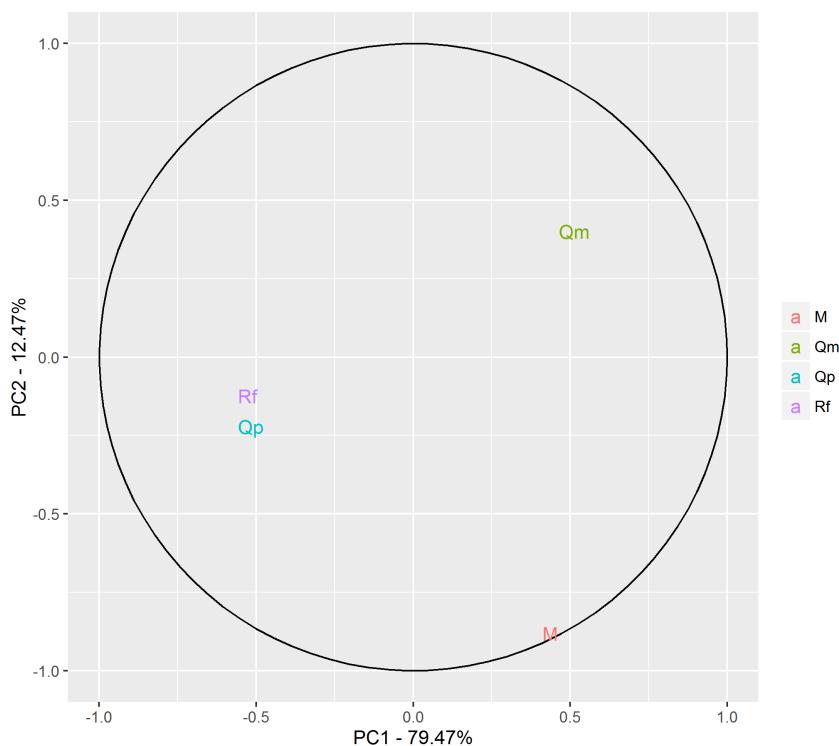
	CP1	CP2	CP3	CP4
Qm	83,53	91,51	93,24	100,00
Rf	88,30	89,06	93,52	100,00
Qp	85,19	87,65	99,68	100,00
M	60,85	99,53	99,53	100,00

**Tabela A.44:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4
Qm	0,91	0,28	0,13	0,26
Rf	-0,94	-0,09	-0,21	0,25
Qp	-0,92	-0,16	0,35	0,06
M	0,78	-0,62	0,00	0,07



**Figura A.21:** Proporção da variância explicada



**Figura A.22:** Coordenadas das variáveis

### A.2.5 *Additive Log-Ratio (Alr)*

**Tabela A.45:** Variâncias das Componentes principais

	CP1	CP2	CP3
Variância	2,37	0,58	0,05
Proporção da Variância	0,79	0,19	0,02
Proporção Acumulada da Variância	0,79	0,98	1,00

**Tabela A.46:** Cargas das Componentes principais

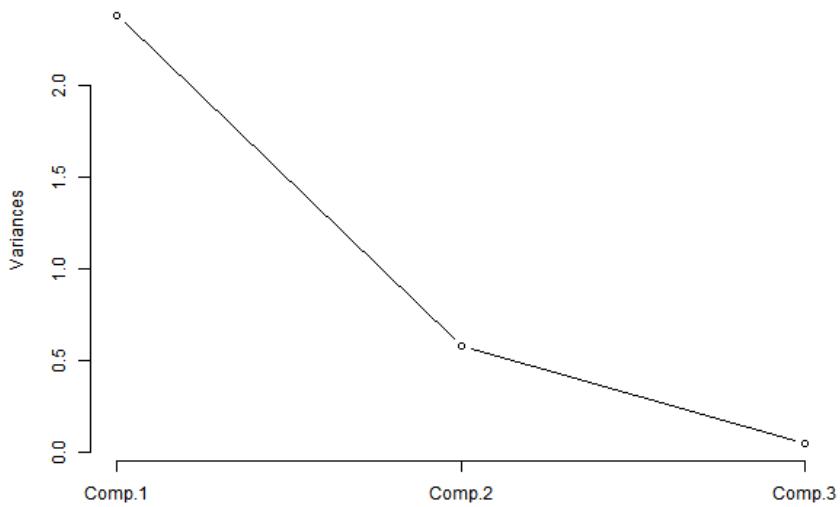
	CP1	CP2	CP3
Qm	0,49	0,87	0,09
Rf	0,63	-0,27	-0,73
Qp	0,61	-0,41	0,68

**Tabela A.47:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

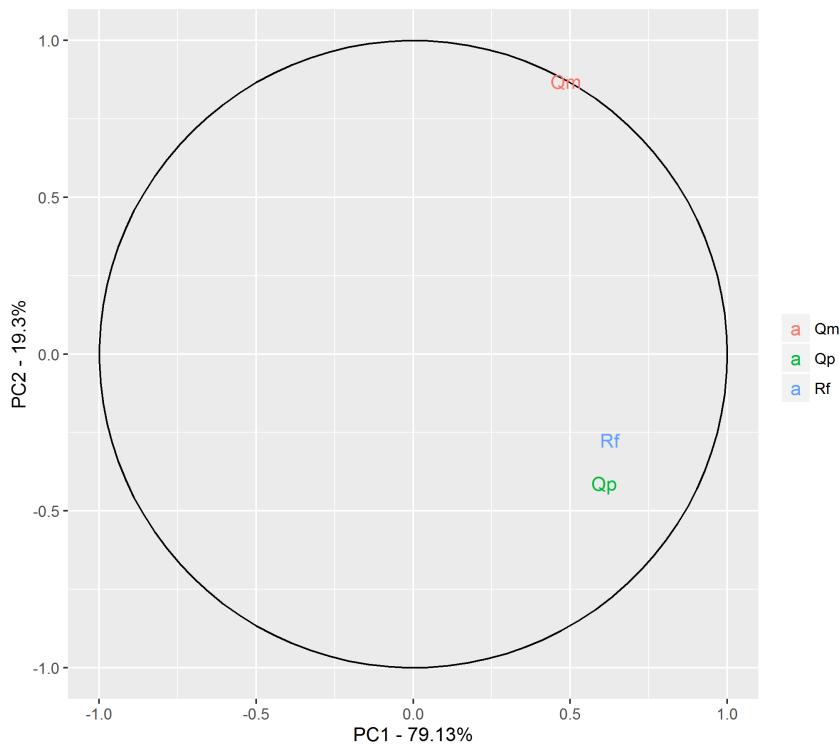
	CP1	CP2	CP3
Qm	56,25	99,96	100,00
Rf	93,14	97,51	100,00
Qp	88,02	97,85	100,00

**Tabela A.48:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3
Qm	0,75	0,66	0,02
Rf	0,97	-0,21	-0,16
Qp	0,94	-0,31	0,15



**Figura A.23:** Proporção da variância explicada



**Figura A.24:** Coordenadas das variáveis

### A.2.6 *Center Log-Ratio* (Clr)

**Tabela A.49:** *Variâncias das Componentes principais*

Variância	3,29	0,48	0,23	0,00
Proporção da Variância	0,82	0,12	0,06	0,00
Proporção Acumulada da Variância	0,82	0,94	1,00	1,00

**Tabela A.50:** *Cargas das Componentes principais*

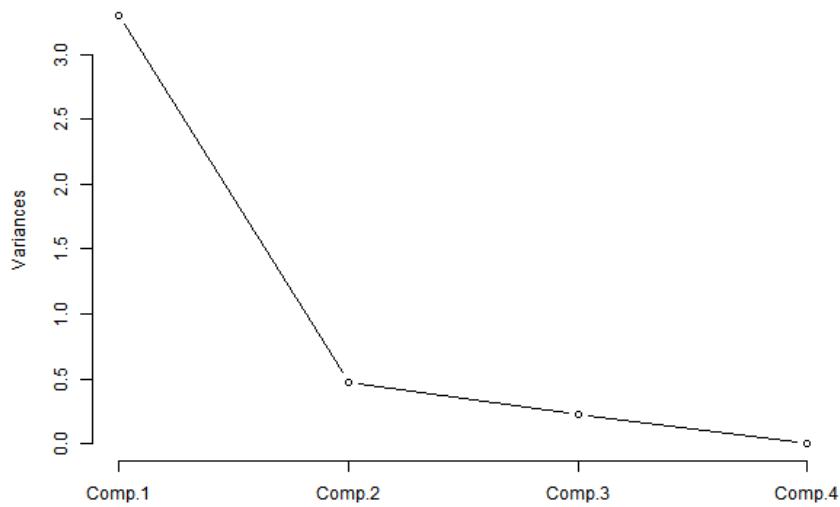
	CP1	CP2	CP3	CP4
Qm	0,47	0,75	-0,15	-0,44
Rf	-0,51	0,24	0,73	-0,38
Qp	-0,52	-0,13	-0,62	-0,57
M	0,50	-0,60	0,24	-0,58

**Tabela A.51:** *Proporção da variância acumulada de cada variável explicada pelas componentes principais*

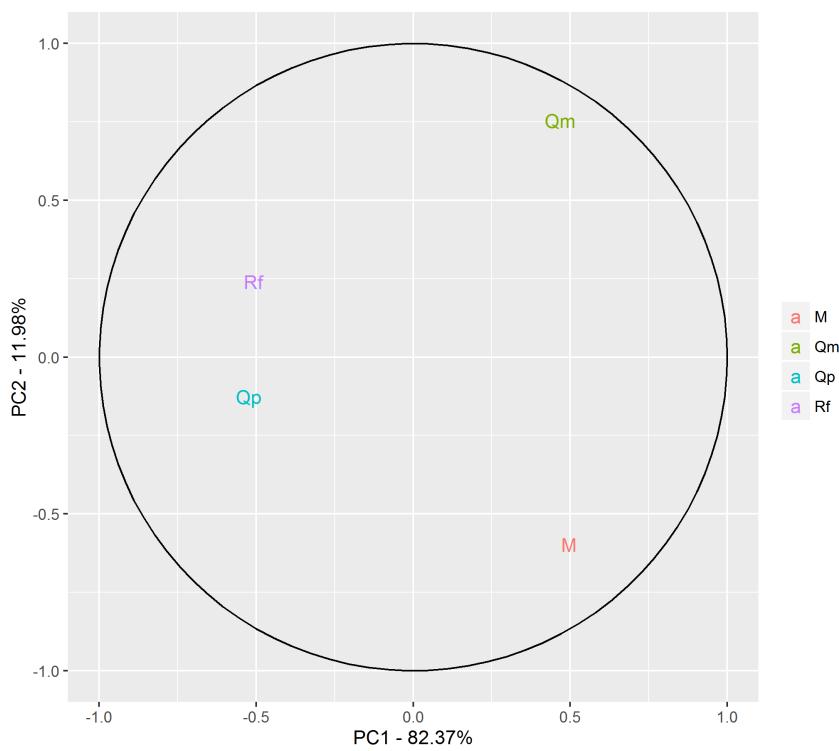
	CP1	CP2	CP3	CP4
Qm	72,22	99,49	100,00	100,00
Rf	85,11	87,87	100,00	100,00
Qp	90,55	91,32	100,00	100,00
M	81,60	98,73	99,99	100,00

**Tabela A.52:** *Matriz de correlação entre as variáveis e as componentes principais*

	CP1	CP2	CP3	CP4
Qm	0,85	0,52	-0,07	-0,00
Rf	-0,92	0,17	0,35	-0,00
Qp	-0,95	-0,09	-0,29	-0,00
M	0,90	-0,41	0,11	-0,00



**Figura A.25:** Proporção da variância explicada



**Figura A.26:** Coordenadas das variáveis

### A.2.7 Transformação Esférica (Esf)

**Tabela A.53:** Variâncias das Componentes principais

	CP1	CP2	CP3
Variância	2,45	0,39	0,16
Proporção da Variância	0,82	0,13	0,05
Proporção Acumulada da Variância	0,82	0,95	1,00

**Tabela A.54:** Cargas das Componentes principais

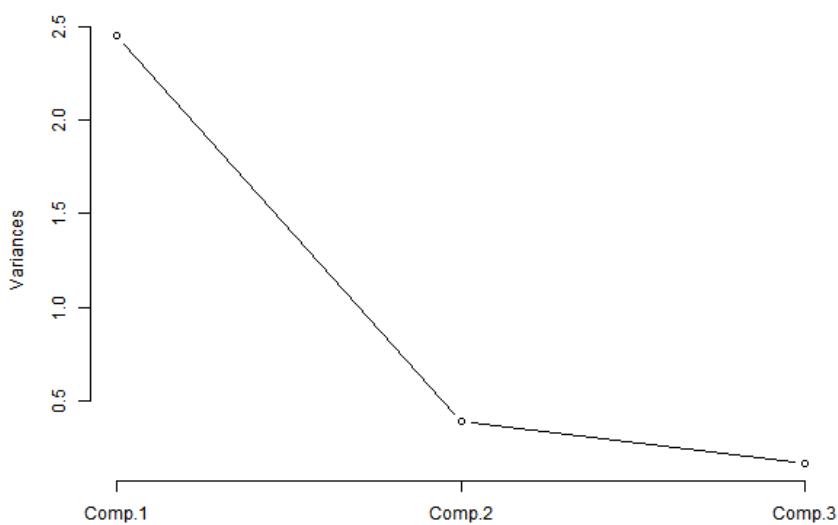
	CP1	CP2	CP3
Qm	0,59	0,46	0,67
Rf	-0,55	0,83	-0,09
Qp	-0,60	-0,31	0,74

**Tabela A.55:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

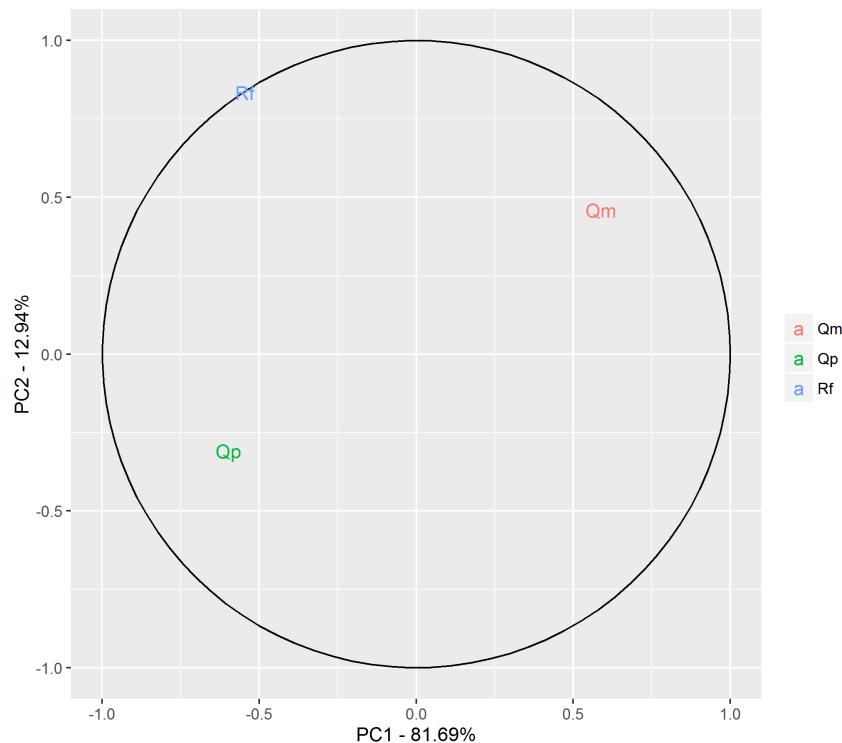
	CP1	CP2	CP3
Qm	84,67	92,85	100,00
Rf	72,92	99,86	100,00
Qp	87,47	91,18	100,00

**Tabela A.56:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3
Qm	0,92	0,29	0,27
Rf	-0,85	0,52	-0,04
Qp	-0,94	-0,19	0,30



**Figura A.27:** Proporção da variância explicada



**Figura A.28:** Coordenadas das variáveis

## A.3 Análise base 3

### A.3.1 Análise Tradicional (Trad)

**Tabela A.57:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
Variância	6,92	1,82	1,42	0,49	0,17	0,08	0,05	0,03	0,01	0,00	0,00
Proporção da Variância	0,63	0,17	0,13	0,04	0,02	0,01	0,00	0,00	0,00	0,00	0,00
Proporção Acumulada da Variância	0,63	0,79	0,92	0,97	0,98	0,99	1,00	1,00	1,00	1,00	1,00

**Tabela A.58:** Cargas das Componentes principais

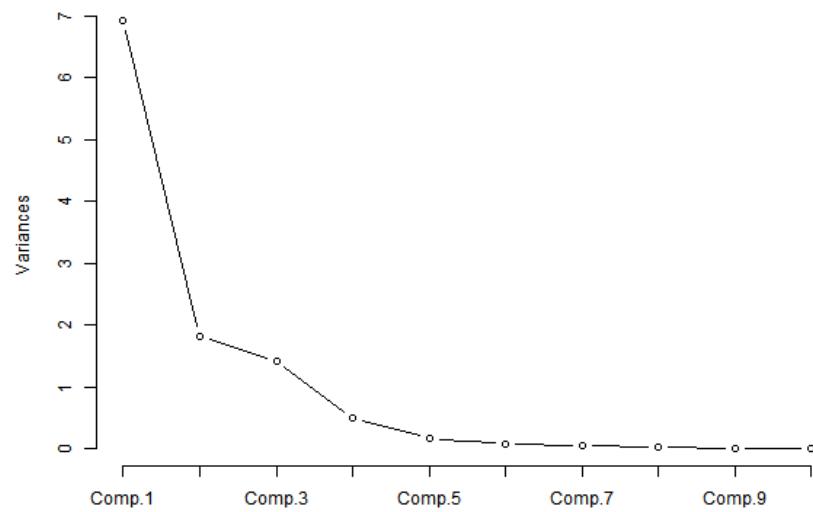
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO</i> <sub>2</sub>	0,36	0,07	0,24	0,04	-0,08	-0,01	-0,07	0,03	-0,10	0,11	0,88
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	-0,30	-0,25	-0,40	0,18	0,13	-0,26	0,25	-0,41	0,46	0,10	0,33
<i>Na</i> <sub>2</sub> <i>O</i>	0,13	-0,51	-0,51	-0,16	0,35	0,20	-0,25	0,05	-0,45	-0,02	0,10
<i>K</i> <sub>2</sub> <i>O</i>	-0,27	0,14	-0,45	0,43	-0,60	0,08	-0,10	0,30	-0,20	0,02	0,11
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	-0,36	0,20	0,08	0,06	0,25	0,18	-0,22	-0,04	0,03	-0,79	0,22
<i>Outros</i>	-0,28	0,15	-0,18	-0,86	-0,31	-0,01	0,05	0,02	0,04	0,02	0,17
<i>CaO</i>	-0,23	-0,54	0,27	-0,03	-0,04	0,13	-0,19	0,57	0,44	0,07	0,07
<i>MgO</i>	-0,35	0,24	0,03	0,05	0,44	0,32	0,53	0,31	-0,19	0,30	0,13
<i>TiO</i> <sub>2</sub>	-0,35	-0,15	0,24	0,00	0,08	-0,76	-0,02	0,13	-0,45	-0,02	0,03
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	-0,25	-0,42	0,37	0,04	-0,35	0,38	0,22	-0,46	-0,31	-0,02	0,01
<i>MnO</i>	-0,35	0,22	0,11	0,03	0,16	0,12	-0,66	-0,28	0,00	0,50	0,00

**Tabela A.59:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

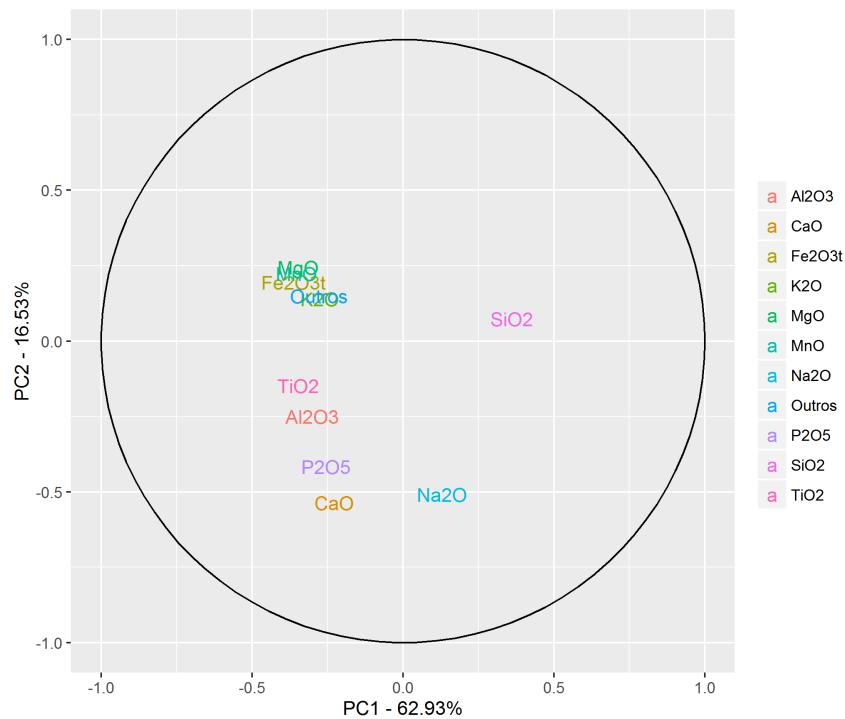
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO</i> <sub>2</sub>	90,26	91,23	99,77	99,85	99,96	99,96	99,99	99,99	100,00	100,00	100,00
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	62,22	73,62	96,32	98,00	98,30	98,87	99,19	99,75	100,00	100,00	100,00
<i>Na</i> <sub>2</sub> <i>O</i>	11,73	58,89	95,76	97,08	99,12	99,44	99,75	99,76	100,00	100,00	100,00
<i>K</i> <sub>2</sub> <i>O</i>	52,15	55,67	84,43	93,66	99,54	99,59	99,64	99,94	99,99	99,99	100,00
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	90,46	97,39	98,21	98,39	99,39	99,65	99,89	99,89	99,89	100,00	100,00
<i>Outros</i>	53,22	57,26	61,94	98,43	99,99	99,99	100,00	100,00	100,00	100,00	100,00
<i>CaO</i>	35,56	87,81	98,31	98,34	98,37	98,50	98,68	99,76	99,99	99,99	100,00
<i>MgO</i>	83,06	93,92	94,08	94,22	97,37	98,21	99,62	99,95	99,99	100,00	100,00
<i>TiO</i> <sub>2</sub>	82,84	86,85	94,94	94,94	95,04	99,70	99,70	99,75	99,99	99,99	100,00
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	44,68	76,18	95,67	95,76	97,75	98,93	99,18	99,89	100,00	100,00	100,00
<i>MnO</i>	86,05	95,19	96,89	96,94	97,36	97,48	99,70	99,96	99,96	100,00	100,00

**Tabela A.60:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO</i> <sub>2</sub>	0,95	0,10	0,29	0,03	-0,03	-0,00	-0,02	0,01	-0,01	0,00	0,00
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	-0,79	-0,34	-0,48	0,13	0,05	-0,08	0,06	-0,07	0,05	0,00	0,00
<i>Na</i> <sub>2</sub> <i>O</i>	0,34	-0,69	-0,61	-0,11	0,14	0,06	-0,06	0,01	-0,05	-0,00	0,00
<i>K</i> <sub>2</sub> <i>O</i>	-0,72	0,19	-0,54	0,30	-0,24	0,02	-0,02	0,05	-0,02	0,00	0,00
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	-0,95	0,26	0,09	0,04	0,10	0,05	-0,05	-0,01	0,00	-0,03	0,00
<i>Outros</i>	-0,73	0,20	-0,22	-0,60	-0,12	-0,00	0,01	0,00	0,00	0,00	0,00
<i>CaO</i>	-0,60	-0,72	0,32	-0,02	-0,02	0,04	-0,04	0,10	0,05	0,00	0,00
<i>MgO</i>	-0,91	0,33	0,04	0,04	0,18	0,09	0,12	0,06	-0,02	0,01	0,00
<i>TiO</i> <sub>2</sub>	-0,91	-0,20	0,28	0,00	0,03	-0,22	-0,00	0,02	-0,05	-0,00	0,00
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	-0,67	-0,56	0,44	0,03	-0,14	0,11	0,05	-0,08	-0,03	-0,00	0,00
<i>MnO</i>	-0,93	0,30	0,13	0,02	0,06	0,03	-0,15	-0,05	0,00	0,02	0,00



**Figura A.29:** Proporção da variância explicada



**Figura A.30:** Coordenadas das variáveis

### A.3.2 Exclusão de uma Variável (Exc)

**Tabela A.61:** *Variâncias das Componentes principais*

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Variância	6,09	1,72	1,40	0,49	0,16	0,08	0,04	0,02	0,01	0,00
Proporção da Variância	0,61	0,17	0,14	0,05	0,02	0,01	0,00	0,00	0,00	0,00
Proporção Acumulada da Variância	0,61	0,78	0,92	0,97	0,99	0,99	1,00	1,00	1,00	1,00

**Tabela A.62:** *Cargas das Componentes principais*

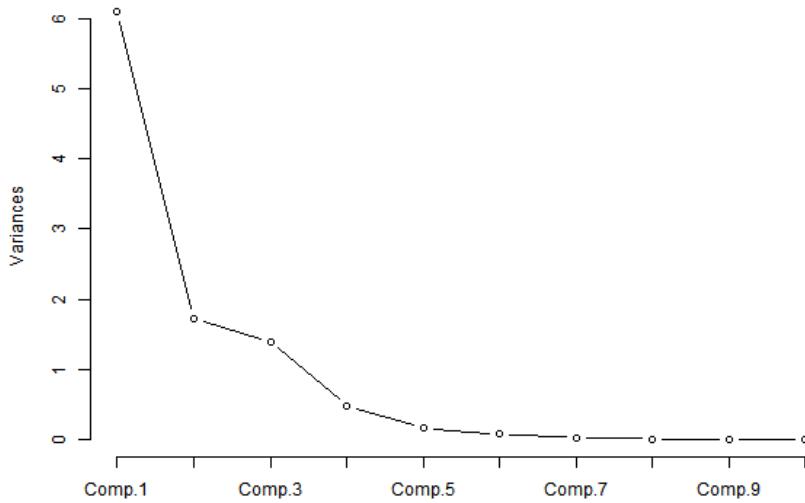
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO</i> <sub>2</sub>	0,39	0,00	0,23	-0,04	0,10	0,03	0,05	-0,08	0,10	0,87
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	-0,33	0,17	-0,42	-0,18	-0,15	0,24	-0,50	-0,06	-0,46	0,33
<i>Na</i> <sub>2</sub> <i>O</i>	0,12	0,47	-0,60	0,18	-0,30	-0,17	0,17	0,13	0,45	0,10
<i>K</i> <sub>2</sub> <i>O</i>	-0,29	-0,23	-0,41	-0,44	0,59	-0,06	0,30	-0,09	0,21	0,11
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	-0,38	-0,23	0,13	-0,07	-0,26	-0,18	0,19	0,77	-0,06	0,23
<i>Outros</i>	-0,29	-0,21	-0,13	0,86	0,29	0,01	-0,01	-0,06	-0,03	0,17
<i>CaO</i>	-0,26	0,55	0,20	0,03	0,05	-0,12	0,58	-0,20	-0,44	0,07
<i>MgO</i>	-0,36	-0,28	0,10	-0,06	-0,49	-0,40	-0,00	-0,57	0,21	0,12
<i>TiO</i> <sub>2</sub>	-0,37	0,14	0,24	-0,01	-0,11	0,74	0,10	-0,04	0,45	0,03
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	-0,28	0,44	0,32	-0,04	0,35	-0,39	-0,50	0,10	0,31	0,01

**Tabela A.63:** *Proporção da variância acumulada de cada variável explicada pelas componentes principais*

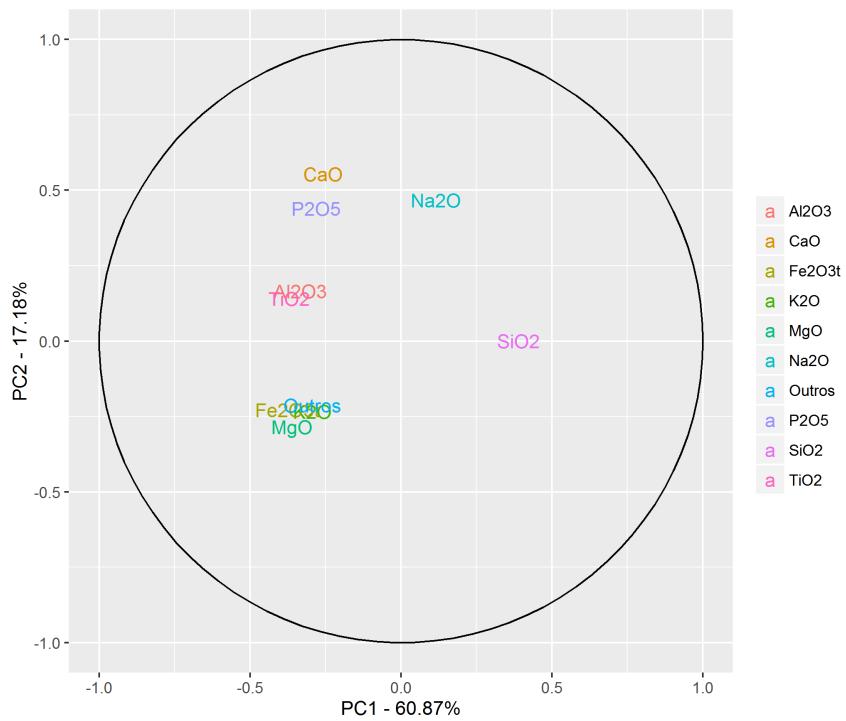
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO</i> <sub>2</sub>	92,39	92,39	99,75	99,82	99,97	99,98	99,99	100,00	100,00	100,00
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	67,27	72,04	96,36	98,02	98,38	98,86	99,76	99,77	100,00	100,00
<i>Na</i> <sub>2</sub> <i>O</i>	8,13	45,72	96,40	97,93	99,40	99,63	99,73	99,76	99,99	100,00
<i>K</i> <sub>2</sub> <i>O</i>	52,24	61,62	84,55	93,97	99,57	99,60	99,94	99,95	100,00	100,00
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	85,95	94,88	97,32	97,54	98,67	98,94	99,07	100,00	100,00	100,00
<i>Outros</i>	52,41	60,14	62,59	98,59	100,00	100,00	100,00	100,00	100,00	100,00
<i>CaO</i>	40,00	92,67	98,30	98,34	98,38	98,49	99,70	99,76	99,99	100,00
<i>MgO</i>	78,79	92,73	94,10	94,28	98,17	99,43	99,43	99,94	99,99	100,00
<i>TiO</i> <sub>2</sub>	83,58	86,98	95,04	95,04	95,24	99,71	99,75	99,75	99,99	100,00
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	47,95	81,28	95,74	95,84	97,77	98,97	99,87	99,89	100,00	100,00

**Tabela A.64:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO<sub>2</sub></i>	0,96	0,00	0,27	-0,03	0,04	0,01	0,01	-0,01	0,01	0,00
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,82	0,22	-0,49	-0,13	-0,06	0,07	-0,09	-0,01	-0,05	0,00
<i>Na<sub>2</sub>O</i>	0,29	0,61	-0,71	0,12	-0,12	-0,05	0,03	0,02	0,05	0,00
<i>K<sub>2</sub>O</i>	-0,72	-0,31	-0,48	-0,31	0,24	-0,02	0,06	-0,01	0,02	0,00
<i>Fe<sub>2</sub>O<sub>3</sub></i>	-0,93	-0,30	0,16	-0,05	-0,11	-0,05	0,04	0,10	-0,01	0,00
<i>Outros</i>	-0,72	-0,28	-0,16	0,60	0,12	0,00	-0,00	-0,01	-0,00	0,00
<i>CaO</i>	-0,63	0,73	0,24	0,02	0,02	-0,03	0,11	-0,02	-0,05	0,00
<i>MgO</i>	-0,89	-0,37	0,12	-0,04	-0,20	-0,11	-0,00	-0,07	0,02	0,00
<i>TiO<sub>2</sub></i>	-0,91	0,18	0,28	-0,01	-0,04	0,21	0,02	-0,01	0,05	0,00
<i>P<sub>2</sub>O<sub>5</sub></i>	-0,69	0,58	0,38	-0,03	0,14	-0,11	-0,09	0,01	0,03	0,00



**Figura A.31:** Proporção da variância explicada



**Figura A.32:** Coordenadas das variáveis

### A.3.3 Variável de Referência (Ref)

**Tabela A.65:** *Variâncias das Componentes principais*

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Variância	4,61	2,33	1,51	0,90	0,27	0,16	0,13	0,07	0,02	0,00
Proporção da Variância	0,46	0,23	0,15	0,09	0,03	0,02	0,01	0,01	0,00	0,00
Proporção Acumulada da Variância	0,46	0,69	0,85	0,93	0,96	0,98	0,99	1,00	1,00	1,00

**Tabela A.66:** *Cargas das Componentes principais*

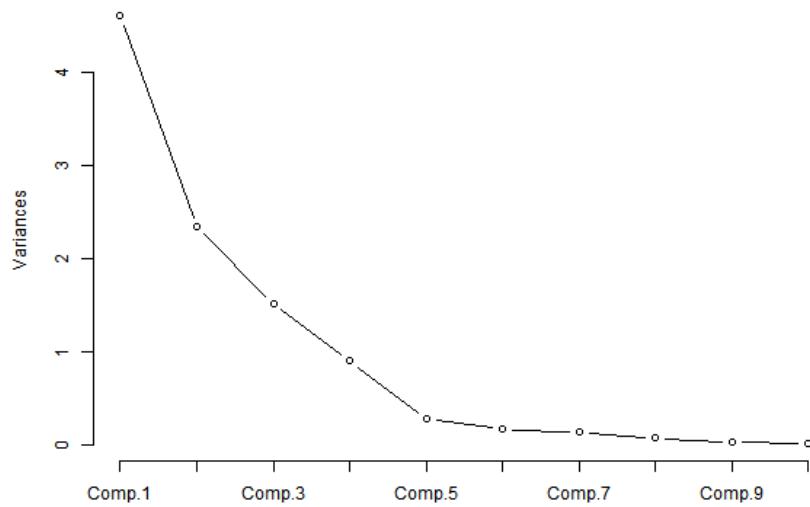
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO<sub>2</sub></i>	-0,42	-0,19	0,02	-0,09	-0,27	0,12	0,66	0,48	0,11	-0,13
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,45	-0,05	-0,16	-0,10	0,00	-0,08	-0,10	-0,31	-0,48	-0,65
<i>Na<sub>2</sub>O</i>	-0,45	-0,07	-0,13	-0,04	-0,06	0,33	-0,06	-0,18	-0,38	0,69
<i>K<sub>2</sub>O</i>	-0,41	-0,17	-0,19	-0,12	0,19	-0,61	-0,01	-0,23	0,51	0,19
<i>Fe<sub>2</sub>O<sub>3</sub></i>	-0,03	0,41	-0,57	-0,28	0,06	-0,14	-0,26	0,57	-0,11	0,04
<i>Outros</i>	-0,13	0,03	-0,29	0,94	0,02	-0,07	-0,01	0,12	-0,01	-0,02
<i>CaO</i>	-0,37	0,35	0,11	0,01	0,13	0,56	-0,29	-0,04	0,53	-0,19
<i>MgO</i>	0,25	0,27	-0,57	-0,06	-0,22	0,18	0,44	-0,48	0,17	-0,03
<i>TiO<sub>2</sub></i>	-0,15	0,50	0,32	0,06	-0,70	-0,33	-0,10	-0,08	-0,01	0,08
<i>P<sub>2</sub>O<sub>5</sub></i>	-0,09	0,56	0,27	0,05	0,57	-0,15	0,45	-0,06	-0,21	0,08

**Tabela A.67:** *Proporção da variância acumulada de cada variável explicada pelas componentes principais*

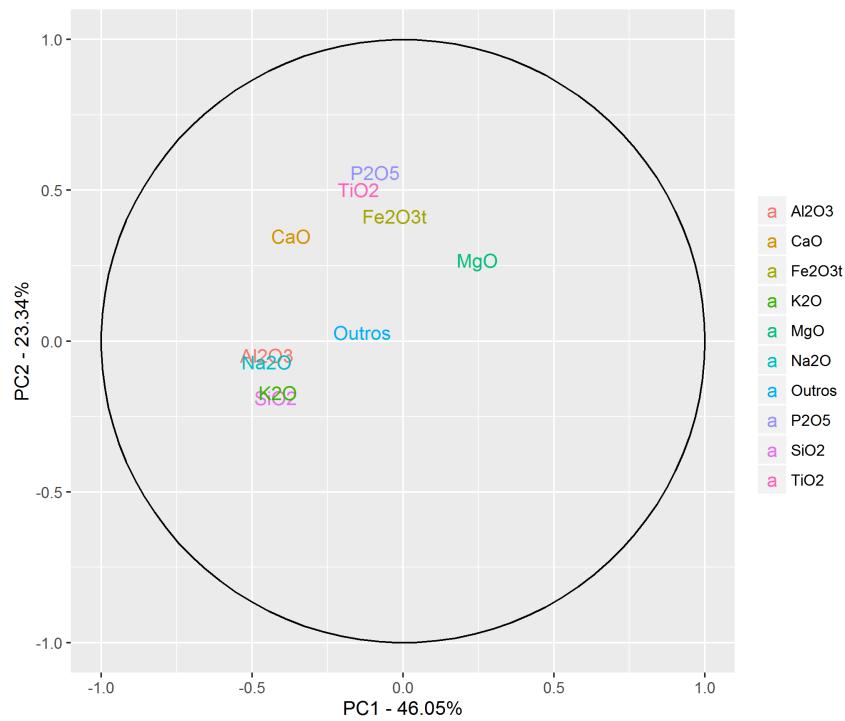
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO<sub>2</sub></i>	81,59	89,79	89,87	90,59	92,54	92,77	98,40	99,97	100,00	100,00
<i>Al<sub>2</sub>O<sub>3</sub></i>	93,34	93,88	97,61	98,48	98,48	98,59	98,73	99,40	99,88	100,00
<i>Na<sub>2</sub>O</i>	93,50	94,61	97,30	97,46	97,56	99,27	99,32	99,55	99,85	100,00
<i>K<sub>2</sub>O</i>	78,70	85,58	90,99	92,31	93,33	99,08	99,08	99,45	99,99	100,00
<i>Fe<sub>2</sub>O<sub>3</sub></i>	0,34	40,31	89,48	96,53	96,63	96,94	97,79	99,97	100,00	100,00
<i>Outros</i>	8,34	8,52	21,25	99,82	99,83	99,90	99,90	100,00	100,00	100,00
<i>CaO</i>	62,80	90,96	92,88	92,89	93,37	98,31	99,39	99,40	99,98	100,00
<i>MgO</i>	28,02	44,78	93,60	93,95	95,31	95,85	98,35	99,94	100,00	100,00
<i>TiO<sub>2</sub></i>	9,99	68,96	84,36	84,71	98,14	99,83	99,96	100,00	100,00	100,00
<i>P<sub>2</sub>O<sub>5</sub></i>	3,91	76,53	87,73	87,93	96,92	97,25	99,87	99,89	99,98	100,00

**Tabela A.68:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO<sub>2</sub></i>	-0,90	-0,29	0,03	-0,08	-0,14	0,05	0,24	0,13	0,02	-0,01
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,97	-0,07	-0,19	-0,09	0,00	-0,03	-0,04	-0,08	-0,07	-0,04
<i>Na<sub>2</sub>O</i>	-0,97	-0,11	-0,16	-0,04	-0,03	0,13	-0,02	-0,05	-0,05	0,04
<i>K<sub>2</sub>O</i>	-0,89	-0,26	-0,23	-0,11	0,10	-0,24	-0,00	-0,06	0,07	0,01
<i>Fe<sub>2</sub>O<sub>3</sub></i>	-0,06	0,63	-0,70	-0,27	0,03	-0,06	-0,09	0,15	-0,02	0,00
<i>Outros</i>	-0,29	0,04	-0,36	0,89	0,01	-0,03	-0,00	0,03	-0,00	-0,00
<i>CaO</i>	-0,79	0,53	0,14	0,01	0,07	0,22	-0,10	-0,01	0,08	-0,01
<i>MgO</i>	0,53	0,41	-0,70	-0,06	-0,12	0,07	0,16	-0,13	0,02	-0,00
<i>TiO<sub>2</sub></i>	-0,32	0,77	0,39	0,06	-0,37	-0,13	-0,04	-0,02	-0,00	0,00
<i>P<sub>2</sub>O<sub>5</sub></i>	-0,20	0,85	0,33	0,04	0,30	-0,06	0,16	-0,01	-0,03	0,00



**Figura A.33:** Proporção da variância explicada



**Figura A.34:** Coordenadas das variáveis

### A.3.4 Transformação Log (Log)

**Tabela A.69:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
Variância	6,96	1,74	1,37	0,53	0,16	0,10	0,06	0,04	0,02	0,01	0,00
Proporção da Variância	0,63	0,16	0,12	0,05	0,01	0,01	0,01	0,00	0,00	0,00	0,00
Proporção Acumulada da Variância	0,63	0,79	0,92	0,96	0,98	0,99	0,99	1,00	1,00	1,00	1,00

**Tabela A.70:** Cargas das Componentes principais

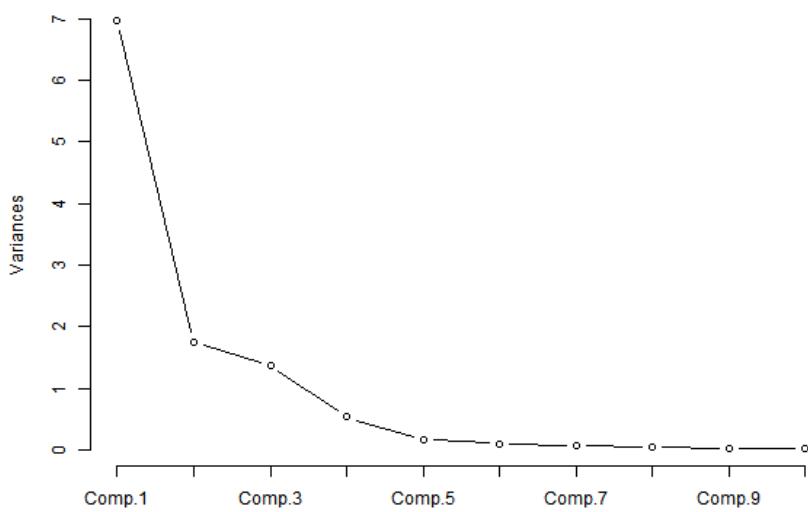
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO<sub>2</sub></i>	0,35	0,21	0,16	-0,03	0,18	-0,52	-0,35	0,38	-0,13	0,46	-0,08
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,28	-0,49	-0,01	0,19	-0,11	0,23	0,16	0,22	0,08	0,70	-0,07
<i>Na<sub>2</sub>O</i>	0,12	-0,61	0,38	-0,04	-0,38	-0,23	-0,30	0,13	0,21	-0,33	0,08
<i>K<sub>2</sub>O</i>	-0,23	-0,39	-0,42	0,36	0,51	-0,27	-0,21	0,04	-0,21	-0,25	-0,08
<i>Fe<sub>2</sub>O<sub>3</sub></i>	-0,36	0,16	-0,07	0,12	-0,21	-0,28	-0,02	0,02	-0,08	0,09	0,83
<i>Outros</i>	-0,24	-0,19	-0,27	-0,89	0,08	-0,08	-0,10	0,00	-0,05	0,08	0,01
<i>CaO</i>	-0,28	-0,07	0,56	-0,02	0,15	-0,15	-0,05	-0,61	-0,39	0,15	-0,10
<i>MgO</i>	-0,35	0,17	-0,09	0,04	-0,51	-0,35	0,24	0,24	-0,31	-0,15	-0,47
<i>TiO<sub>2</sub></i>	-0,34	0,19	0,22	0,01	0,02	0,52	-0,56	0,38	-0,23	-0,13	-0,02
<i>P<sub>2</sub>O<sub>5</sub></i>	-0,31	0,07	0,43	-0,09	0,46	-0,14	0,43	0,38	0,34	-0,18	0,02
<i>MnO</i>	-0,35	0,23	-0,10	0,10	-0,07	-0,17	-0,39	-0,28	0,68	0,11	-0,24

**Tabela A.71:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

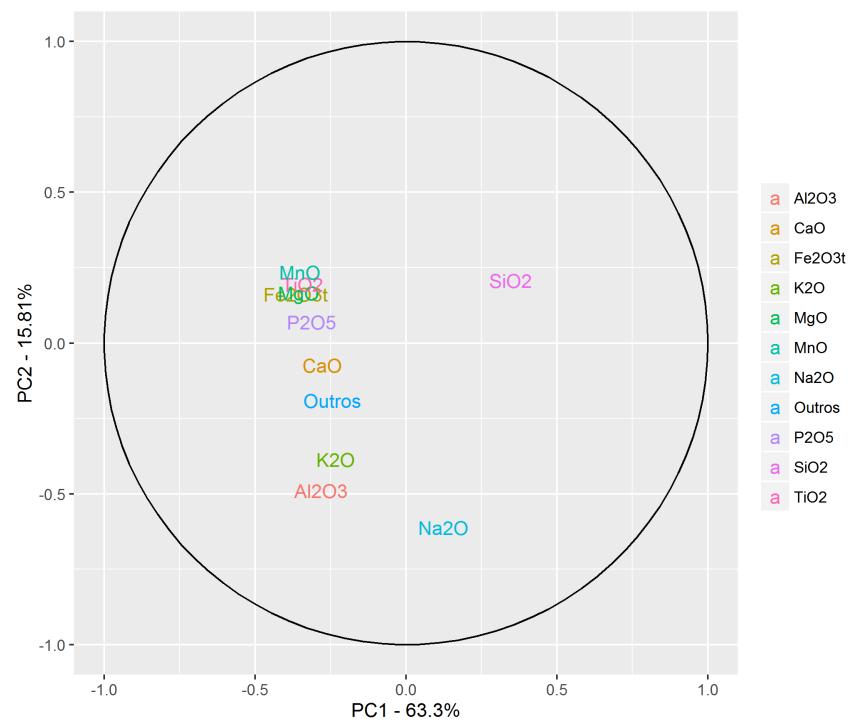
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO<sub>2</sub></i>	84,05	91,50	95,15	95,21	95,74	98,39	99,17	99,69	99,72	99,99	100,00
<i>Al<sub>2</sub>O<sub>3</sub></i>	54,63	96,37	96,38	98,31	98,51	99,02	99,19	99,36	99,37	100,00	100,00
<i>Na<sub>2</sub>O</i>	10,86	76,07	96,23	96,31	98,63	99,15	99,71	99,77	99,86	100,00	100,00
<i>K<sub>2</sub>O</i>	37,46	63,34	87,68	94,57	98,83	99,53	99,81	99,82	99,91	99,99	100,00
<i>Fe<sub>2</sub>O<sub>3</sub></i>	92,24	96,79	97,53	98,28	99,02	99,77	99,77	99,77	99,78	99,79	99,99
<i>Outros</i>	41,02	47,32	57,67	99,74	99,86	99,93	100,00	100,00	100,00	100,00	100,00
<i>CaO</i>	53,06	54,01	97,73	97,75	98,10	98,31	98,32	99,67	99,97	100,00	100,00
<i>MgO</i>	87,69	92,45	93,53	93,62	97,94	99,14	99,50	99,71	99,90	99,93	100,00
<i>TiO<sub>2</sub></i>	81,48	88,09	94,72	94,72	94,73	97,39	99,34	99,86	99,97	99,99	100,00
<i>P<sub>2</sub>O<sub>5</sub></i>	67,86	68,72	94,02	94,42	97,84	98,04	99,20	99,73	99,96	100,00	100,00
<i>MnO</i>	85,93	95,53	96,90	97,44	97,51	97,80	98,77	99,05	99,97	99,99	100,00

**Tabela A.72:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO</i> <sub>2</sub>	0,92	0,27	0,19	-0,02	0,07	-0,16	-0,09	0,07	-0,02	0,05	-0,00
<i>Al</i> <sub>2</sub> O <sub>3</sub>	-0,74	-0,65	-0,01	0,14	-0,04	0,07	0,04	0,04	0,01	0,08	-0,00
<i>Na</i> <sub>2</sub> O	0,33	-0,81	0,45	-0,03	-0,15	-0,07	-0,08	0,02	0,03	-0,04	0,00
<i>K</i> <sub>2</sub> O	-0,61	-0,51	-0,49	0,26	0,21	-0,08	-0,05	0,01	-0,03	-0,03	-0,00
<i>Fe</i> <sub>2</sub> O <sub>3</sub>	-0,96	0,21	-0,09	0,09	-0,09	-0,09	-0,01	0,00	-0,01	0,01	0,05
<i>Outros</i>	-0,64	-0,25	-0,32	-0,65	0,03	-0,03	-0,03	0,00	-0,01	0,01	0,00
<i>Ca</i> O	-0,73	-0,10	0,66	-0,02	0,06	-0,05	-0,01	-0,12	-0,05	0,02	-0,01
<i>Mg</i> O	-0,94	0,22	-0,10	0,03	-0,21	-0,11	0,06	0,05	-0,04	-0,02	-0,03
<i>Ti</i> O <sub>2</sub>	-0,90	0,26	0,26	0,00	0,01	0,16	-0,14	0,07	-0,03	-0,01	-0,00
<i>P</i> <sub>2</sub> O <sub>5</sub>	-0,82	0,09	0,50	-0,06	0,19	-0,04	0,11	0,07	0,05	-0,02	0,00
MnO	-0,93	0,31	-0,12	0,07	-0,03	-0,05	-0,10	-0,05	0,10	0,01	-0,01



**Figura A.35:** Proporção da variância explicada



**Figura A.36:** Coordenadas das variáveis

### A.3.5 Additive Log-Ratio (Alr)

**Tabela A.73:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Variância	5,13	2,18	1,41	0,80	0,24	0,13	0,06	0,04	0,01	0,00
Proporção da Variância	0,51	0,22	0,14	0,08	0,02	0,01	0,01	0,00	0,00	0,00
Proporção Acumulada da Variância	0,51	0,73	0,87	0,95	0,98	0,99	0,99	1,00	1,00	1,00

**Tabela A.74:** Cargas das Componentes principais

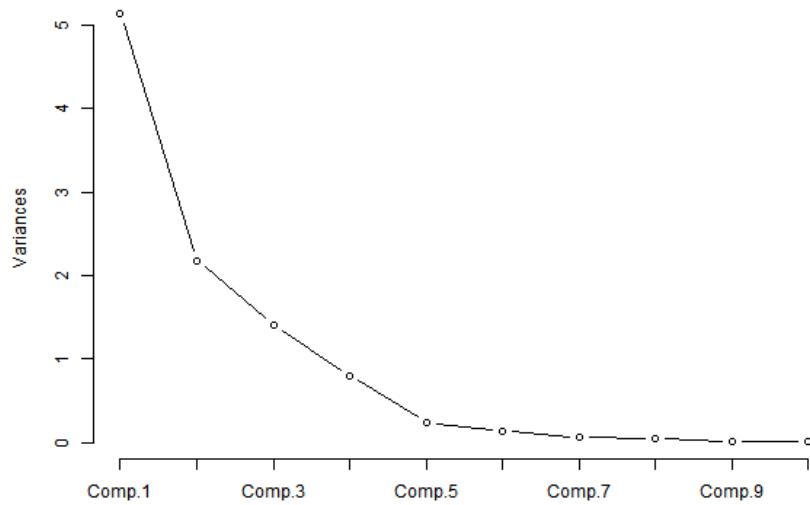
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
$SiO_2$	-0,41	-0,17	-0,07	0,13	0,14	-0,48	0,64	-0,27	0,15	-0,20
$Al_2O_3$	-0,42	-0,02	-0,23	0,09	0,05	0,12	-0,20	0,33	-0,44	-0,64
$Na_2O$	-0,43	-0,05	-0,13	0,08	0,04	-0,26	-0,20	0,01	-0,46	0,69
$K_2O$	-0,39	-0,13	-0,33	0,11	-0,05	0,44	0,10	0,37	0,56	0,24
$Fe_2O_3$	0,03	0,55	-0,36	0,41	-0,03	0,35	0,16	-0,47	-0,15	0,03
<i>Outros</i>	-0,00	0,16	-0,55	-0,81	0,05	-0,05	0,04	-0,14	0,00	-0,00
$CaO$	-0,40	0,23	0,14	-0,01	-0,22	-0,21	-0,59	-0,37	0,42	-0,15
$MgO$	0,26	0,45	-0,30	0,25	0,16	-0,54	-0,08	0,47	0,21	-0,00
$TiO_2$	-0,23	0,38	0,41	-0,17	0,74	0,19	0,05	0,07	0,04	0,05
$P_2O_5$	-0,21	0,47	0,34	-0,20	-0,59	-0,01	0,36	0,29	-0,11	0,05

**Tabela A.75:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

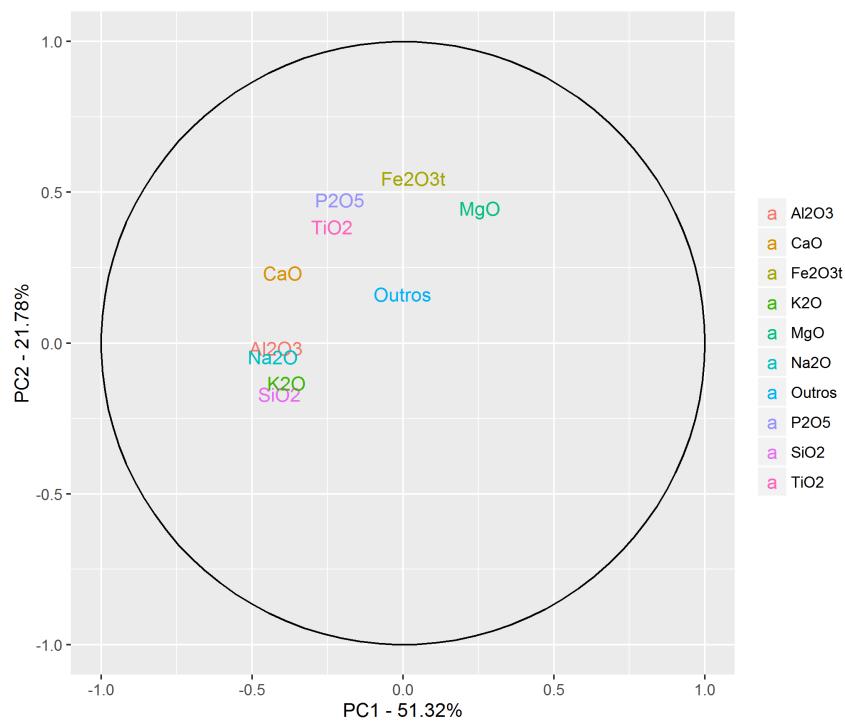
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
$SiO_2$	85,70	92,03	92,66	93,92	94,35	97,41	99,64	99,96	99,98	100,00
$Al_2O_3$	90,42	90,48	98,11	98,75	98,80	98,98	99,19	99,65	99,83	100,00
$Na_2O$	94,92	95,39	97,90	98,47	98,51	99,40	99,62	99,62	99,82	100,00
$K_2O$	76,24	80,10	95,40	96,39	96,44	99,05	99,11	99,69	99,98	100,00
$Fe_2O_3$	0,62	65,44	83,36	97,14	97,16	98,85	99,00	99,97	99,99	100,00
<i>Outros</i>	0,00	5,75	47,67	99,82	99,89	99,92	99,93	100,00	100,00	100,00
$CaO$	81,19	92,89	95,64	95,65	96,78	97,36	99,25	99,83	100,00	100,00
$MgO$	33,50	77,14	89,41	94,46	95,06	98,97	99,01	99,95	99,99	100,00
$TiO_2$	28,27	60,47	84,08	86,43	99,48	99,97	99,98	100,00	100,00	100,00
$P_2O_5$	22,33	71,32	87,38	90,74	98,91	98,91	99,62	99,99	100,00	100,00

**Tabela A.76:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO<sub>2</sub></i>	-0,93	-0,25	-0,08	0,11	0,07	-0,17	0,15	-0,06	0,01	-0,01
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,95	-0,03	-0,28	0,08	0,02	0,04	-0,05	0,07	-0,04	-0,04
<i>Na<sub>2</sub>O</i>	-0,97	-0,07	-0,16	0,08	0,02	-0,09	-0,05	0,00	-0,04	0,04
<i>K<sub>2</sub>O</i>	-0,87	-0,20	-0,39	0,10	-0,02	0,16	0,02	0,08	0,05	0,01
<i>Fe<sub>2</sub>O<sub>3</sub></i>	0,08	0,81	-0,42	0,37	-0,01	0,13	0,04	-0,10	-0,01	0,00
<i>Outros</i>	-0,00	0,24	-0,65	-0,72	0,03	-0,02	0,01	-0,03	0,00	-0,00
<i>CaO</i>	-0,90	0,34	0,17	-0,01	-0,11	-0,08	-0,14	-0,08	0,04	-0,01
<i>MgO</i>	0,58	0,66	-0,35	0,22	0,08	-0,20	-0,02	0,10	0,02	-0,00
<i>TiO<sub>2</sub></i>	-0,53	0,57	0,49	-0,15	0,36	0,07	0,01	0,01	0,00	0,00
<i>P<sub>2</sub>O<sub>5</sub></i>	-0,47	0,70	0,40	-0,18	-0,29	-0,00	0,08	0,06	-0,01	0,00



**Figura A.37:** Proporção da variância explicada



**Figura A.38:** Coordenadas das variáveis

### A.3.6 Center Log-Ratio (Clr)

**Tabela A.77:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
Variância	6,20	2,40	1,48	0,37	0,22	0,18	0,10	0,03	0,01	0,01	0,00
Proporção da Variância	0,56	0,22	0,13	0,03	0,02	0,02	0,01	0,00	0,00	0,00	0,00
Proporção Acumulada da Variância	0,56	0,78	0,92	0,95	0,97	0,99	1,00	1,00	1,00	1,00	1,00

**Tabela A.78:** Cargas das Componentes principais

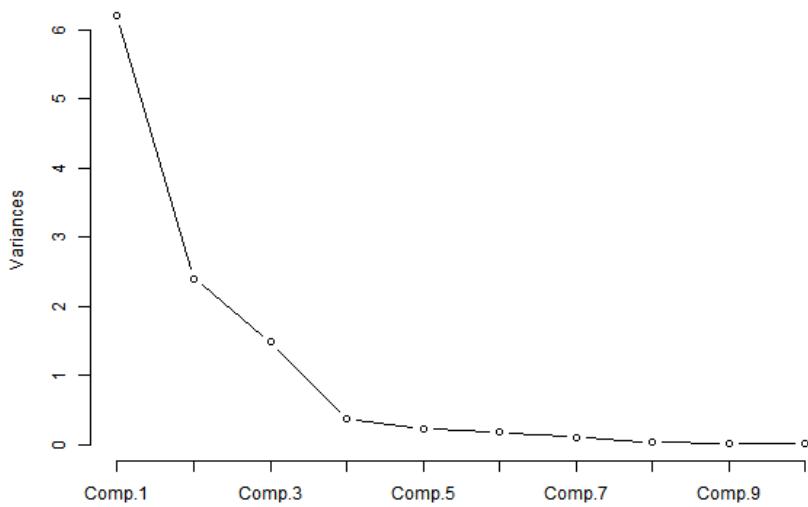
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO</i> <sub>2</sub>	-0,35	0,07	-0,26	-0,41	-0,47	0,27	0,30	0,25	-0,20	0,22	-0,32
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	-0,38	-0,02	-0,19	0,23	0,28	-0,30	0,05	-0,48	-0,46	0,36	-0,18
<i>Na</i> <sub>2</sub> <i>O</i>	-0,39	0,09	-0,06	0,01	-0,21	-0,19	-0,06	-0,33	0,28	-0,68	-0,33
<i>K</i> <sub>2</sub> <i>O</i>	-0,32	-0,21	-0,32	0,22	0,55	0,19	0,04	0,42	0,37	0,01	-0,21
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	0,35	-0,05	-0,36	0,28	-0,03	-0,04	-0,03	0,31	-0,58	-0,44	-0,19
<i>Outros</i>	0,11	-0,42	0,57	-0,16	0,10	0,00	-0,17	0,06	-0,12	0,04	-0,63
<i>Ca</i> <i>O</i>	-0,16	0,56	0,11	0,23	-0,13	-0,08	-0,63	0,30	0,00	0,22	-0,18
<i>Mg</i> <i>O</i>	0,36	-0,10	-0,19	0,36	-0,36	-0,37	0,24	-0,02	0,41	0,32	-0,32
<i>Ti</i> <i>O</i> <sub>2</sub>	0,25	0,41	-0,14	-0,57	0,41	-0,45	0,13	0,10	0,06	-0,02	-0,18
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	0,16	0,52	0,26	0,29	0,17	0,47	0,44	-0,19	-0,00	-0,05	-0,27
<i>Mn</i> <i>O</i>	0,32	-0,05	-0,45	-0,20	0,04	0,44	-0,46	-0,44	0,12	0,10	-0,19

**Tabela A.79:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

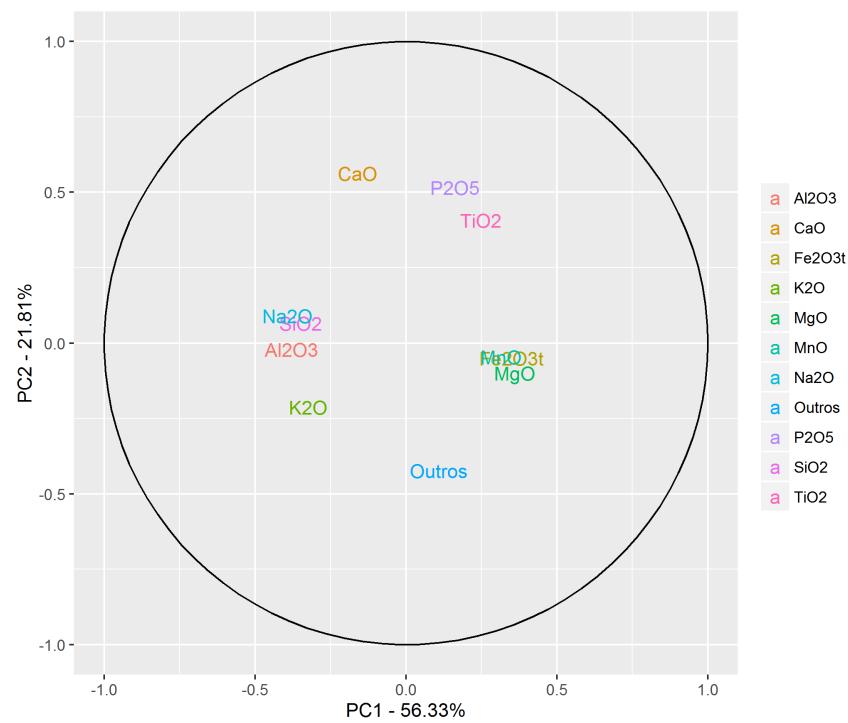
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO</i> <sub>2</sub>	75,12	76,15	86,41	92,57	97,41	98,78	99,71	99,91	99,96	100,00	100,00
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	88,45	88,56	93,68	95,55	97,23	98,88	98,91	99,62	99,88	100,00	100,00
<i>Na</i> <sub>2</sub> <i>O</i>	94,99	96,93	97,45	97,45	98,42	99,11	99,14	99,47	99,56	100,00	100,00
<i>K</i> <sub>2</sub> <i>O</i>	64,47	75,28	90,05	91,80	98,57	99,26	99,28	99,82	99,99	99,99	100,00
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	76,07	76,68	96,09	99,06	99,08	99,11	99,12	99,42	99,82	100,00	100,00
<i>Outros</i>	7,39	50,47	98,47	99,45	99,68	99,68	99,97	99,98	100,00	100,00	100,00
<i>Ca</i> <i>O</i>	15,63	91,49	93,17	95,18	95,58	95,70	99,67	99,96	99,96	100,00	100,00
<i>Mg</i> <i>O</i>	81,20	83,58	89,13	93,76	96,58	99,13	99,71	99,71	99,91	100,00	100,00
<i>Ti</i> <i>O</i> <sub>2</sub>	38,08	77,68	80,59	92,41	96,12	99,81	99,97	100,00	100,00	100,00	100,00
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	16,60	80,54	90,16	93,16	93,84	97,94	99,88	100,00	100,00	100,00	100,00
<i>Mn</i> <i>O</i>	61,68	62,18	92,25	93,74	93,78	97,29	99,38	99,98	100,00	100,00	100,00

**Tabela A.80:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
<i>SiO<sub>2</sub></i>	-0,87	0,10	-0,32	-0,25	-0,22	0,12	0,10	0,04	-0,02	0,02	-0,00
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,94	-0,03	-0,23	0,14	0,13	-0,13	0,02	-0,08	-0,05	0,03	-0,00
<i>Na<sub>2</sub>O</i>	-0,97	0,14	-0,07	0,00	-0,10	-0,08	-0,02	-0,06	0,03	-0,07	-0,00
<i>K<sub>2</sub>O</i>	-0,80	-0,33	-0,38	0,13	0,26	0,08	0,01	0,07	0,04	0,00	-0,00
<i>Fe<sub>2</sub>O<sub>3</sub></i>	0,87	-0,08	-0,44	0,17	-0,01	-0,02	-0,01	0,05	-0,06	-0,04	-0,00
<i>Outros</i>	0,27	-0,66	0,69	-0,10	0,05	0,00	-0,05	0,01	-0,01	0,00	-0,00
<i>CaO</i>	-0,40	0,87	0,13	0,14	-0,06	-0,03	-0,20	0,05	0,00	0,02	-0,00
<i>MgO</i>	0,90	-0,15	-0,24	0,22	-0,17	-0,16	0,08	-0,00	0,04	0,03	-0,00
<i>TiO<sub>2</sub></i>	0,62	0,63	-0,17	-0,34	0,19	-0,19	0,04	0,02	0,01	-0,00	-0,00
<i>P<sub>2</sub>O<sub>5</sub></i>	0,41	0,80	0,31	0,17	0,08	0,20	0,14	-0,03	-0,00	-0,00	-0,00
<i>MnO</i>	0,79	-0,07	-0,55	-0,12	0,02	0,19	-0,14	-0,08	0,01	0,01	-0,00



**Figura A.39:** Proporção da variância explicada



**Figura A.40:** Coordenadas das variáveis

### A.3.7 Transformação Esférica (Esf)

**Tabela A.81:** Variâncias das Componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Variância	4,85	2,12	1,45	0,69	0,43	0,23	0,11	0,05	0,04	0,01
Proporção da Variância	0,49	0,21	0,15	0,07	0,04	0,02	0,01	0,01	0,00	0,00
Proporção Acumulada da Variância	0,49	0,70	0,84	0,91	0,95	0,98	0,99	0,99	1,00	1,00

**Tabela A.82:** Cargas das Componentes principais

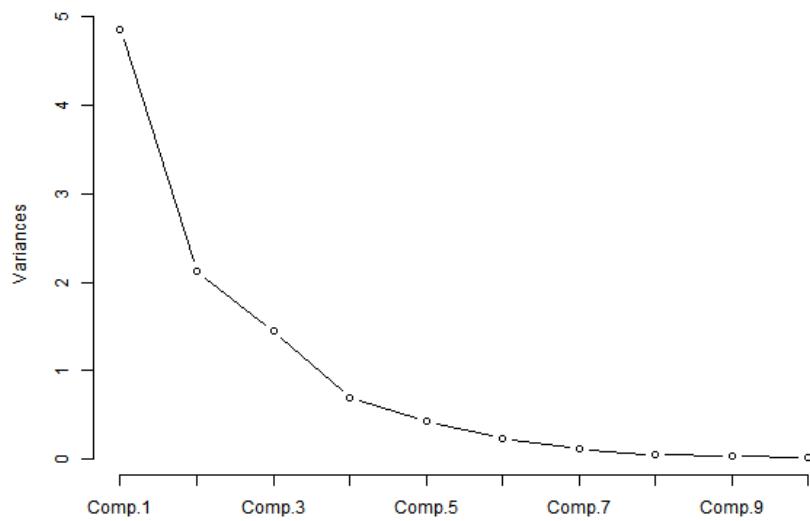
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO</i> <sub>2</sub>	-0,35	-0,21	-0,03	0,07	0,83	-0,07	-0,28	0,15	-0,03	-0,17
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	-0,42	-0,11	0,00	-0,18	-0,35	-0,12	-0,46	-0,49	0,02	-0,43
<i>Na</i> <sub>2</sub> <i>O</i>	-0,43	-0,04	-0,10	-0,11	-0,01	-0,45	0,08	-0,18	0,10	0,73
<i>K</i> <sub>2</sub> <i>O</i>	-0,35	-0,24	-0,19	-0,46	-0,22	0,39	0,02	0,48	-0,37	0,08
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	0,11	-0,54	0,44	-0,08	-0,07	0,24	-0,21	0,10	0,58	0,20
<i>Outros</i>	0,16	0,35	-0,61	-0,23	0,07	0,21	-0,41	0,04	0,44	0,13
<i>CaO</i>	-0,42	0,19	-0,02	-0,05	-0,04	-0,03	0,55	0,19	0,56	-0,36
<i>MgO</i>	0,36	-0,18	-0,06	-0,49	-0,02	-0,68	-0,02	0,29	0,02	-0,22
<i>TiO</i> <sub>2</sub>	-0,04	-0,40	-0,47	0,64	-0,29	-0,19	-0,08	0,27	0,06	-0,06
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	-0,18	0,49	0,41	0,18	-0,20	-0,18	-0,43	0,51	-0,01	0,06

**Tabela A.83:** Proporção da variância acumulada de cada variável explicada pelas componentes principais

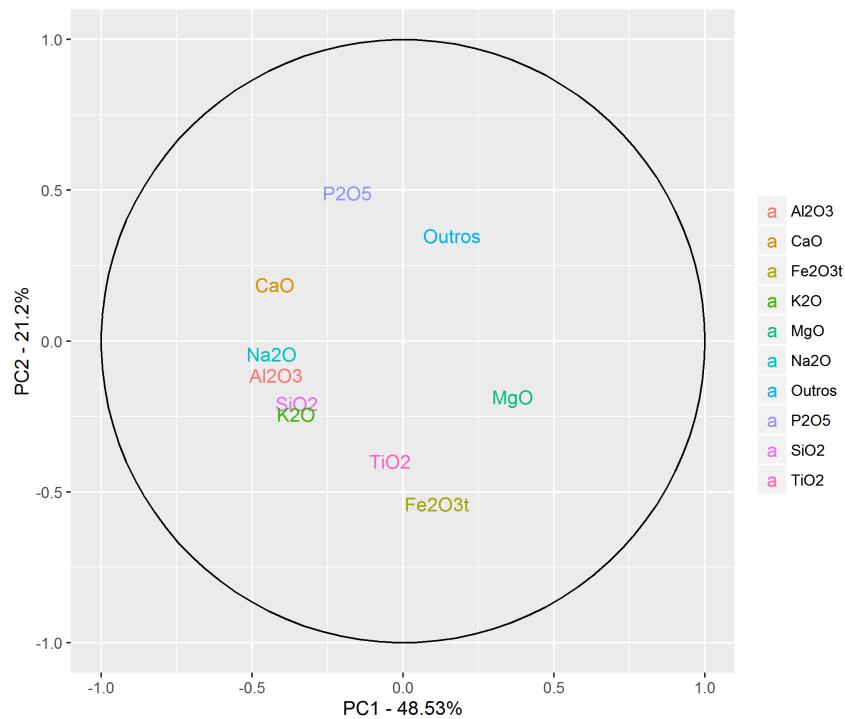
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO</i> <sub>2</sub>	59,58	68,67	68,76	69,11	98,81	98,94	99,83	99,94	99,94	100,00
<i>Al</i> <sub>2</sub> <i>O</i> <sub>3</sub>	85,33	88,08	88,08	90,38	95,74	96,07	98,48	99,73	99,73	100,00
<i>Na</i> <sub>2</sub> <i>O</i>	91,61	92,01	93,42	94,29	94,29	98,94	99,02	99,19	99,23	100,00
<i>K</i> <sub>2</sub> <i>O</i>	60,63	73,13	78,23	92,66	94,65	98,28	98,29	99,47	99,99	100,00
<i>Fe</i> <sub>2</sub> <i>O</i> <sub>3</sub>	6,26	68,30	96,06	96,50	96,73	98,05	98,57	98,62	99,94	100,00
<i>Outros</i>	12,83	38,60	92,46	96,04	96,27	97,31	99,21	99,22	99,98	100,00
<i>CaO</i>	87,19	94,61	94,68	94,89	94,95	94,98	98,42	98,61	99,82	100,00
<i>MgO</i>	64,48	71,71	72,22	88,73	88,74	99,49	99,49	99,93	99,93	100,00
<i>TiO</i> <sub>2</sub>	0,88	34,48	66,71	95,09	98,70	99,51	99,59	99,98	99,99	100,00
<i>P</i> <sub>2</sub> <i>O</i> <sub>5</sub>	16,53	67,74	91,74	94,02	95,70	96,50	98,64	99,99	99,99	100,00

**Tabela A.84:** Matriz de correlação entre as variáveis e as componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<i>SiO<sub>2</sub></i>	-0,77	-0,30	-0,03	0,06	0,55	-0,04	-0,09	0,03	-0,01	-0,02
<i>Al<sub>2</sub>O<sub>3</sub></i>	-0,92	-0,17	0,00	-0,15	-0,23	-0,06	-0,16	-0,11	0,00	-0,05
<i>Na<sub>2</sub>O</i>	-0,96	-0,06	-0,12	-0,09	-0,01	-0,22	0,03	-0,04	0,02	0,09
<i>K<sub>2</sub>O</i>	-0,78	-0,35	-0,23	-0,38	-0,14	0,19	0,01	0,11	-0,07	0,01
<i>Fe<sub>2</sub>O<sub>3</sub></i>	0,25	-0,79	0,53	-0,07	-0,05	0,11	-0,07	0,02	0,11	0,02
<i>Outros</i>	0,36	0,51	-0,73	-0,19	0,05	0,10	-0,14	0,01	0,09	0,01
<i>CaO</i>	-0,93	0,27	-0,03	-0,05	-0,02	-0,02	0,19	0,04	0,11	-0,04
<i>MgO</i>	0,80	-0,27	-0,07	-0,41	-0,01	-0,33	-0,01	0,07	0,00	-0,03
<i>TiO<sub>2</sub></i>	-0,09	-0,58	-0,57	0,53	-0,19	-0,09	-0,03	0,06	0,01	-0,01
<i>P<sub>2</sub>O<sub>5</sub></i>	-0,41	0,72	0,49	0,15	-0,13	-0,09	-0,15	0,12	-0,00	0,01



**Figura A.41:** Proporção da variância explicada



**Figura A.42:** Coordenadas das variáveis

# Apêndice B

## Funções R

Esta seção contém todas as funções que foram desenvolvidas para realizar as análises desta dissertação.

```
#####
##### Lista de pacotes necessarios
#####
library(MASS)
library(xtable)
library(graphics)
library(car)
library(compositions)
library(ggplot2)
library(reshape2)
#####

#####
##### Funcao que plota as coordenadas a primeira e segundas
##### componentes da analise de componentes principais
##### fit objeto da classe princomp
#####

Comp_Graf = function(fit){
  theta <- seq(0,2*pi,length.out = 100)
  circle <- data.frame(x = cos(theta), y = sin(theta))
  Prop_exp = (fit$sdev)^2/sum(fit$sdev^2)
  Cargas = loadings(fit)[]
  p <- ggplot(circle,aes(x,y)) + geom_path()
  loadings <- data.frame(Cargas,
    .names = row.names(Cargas))
  fim = p + geom_text(data=loadings,
    mapping=aes(x = Comp.1, y = Comp.2, label = .names,
    colour = .names)) +
  coord_fixed(ratio=1) +
  labs(x = paste("PC1 - ",round(Prop_exp[1]*100,2),"%",sep = ""),
    y = paste("PC2 - ",round(Prop_exp[2]*100,2),"%",sep = ""))
}
```

```

theme(legend.title=element_blank())
return(fim)
}
#####
##### Funcao que plota os individuos com os peso das
##### primeira e segundas componentes da
##### analise de componentes principais
##### fit objeto da classe princomp
##### group Variavel com a separacao dos grupos da amostra
#####
Comp_Graf_biplot = function(fit ,group){
  g <- ggbiplot(fit , obs.scale = 1, var.scale = 1,
                 groups = group , ellipse = TRUE,
                 circle = TRUE)
  g <- g + scale_color_discrete(name = ' ')
  g <- g + theme(legend.direction = 'horizontal',
                  legend.position = 'top')
  print(g)
}

#####
##### Funcao que plota todos os graficos das componentes principais
##### NOME String com o nome de saida do arquivo dos graficos
##### fit obejto da classe princomp
##### group Variavel com a separacao dos grupos da amostra
#####
PLOT_Comp_Princ = function(NOME,FIT ,grupo = FALSE){

  png(paste("Var",NOME,".png",sep = "") ,width = 600, height = 450)
  plot(FIT,type="lines" , main = "")
  dev.off()
  Comp_Graf(FIT)
  ggsave(paste("Coord",NOME,".png",sep = ""))
  if (grupo != FALSE){
    Comp_Graf_biplot(FIT ,grupo)
    ggsave(paste("Bivar",NOME,".png",sep = ""))
  }
}

#####
##### Funcao que ajusta a tabela da explicacao da variancia
##### de cada variavel dentro de cada componente para impressao
##### fit objeto da classe princomp
#####

print_comp_var = function( fit ){
  Cargas = loadings(fit) []
  perc_da_var = (((fit$sdev)^2)*t((Cargas^2)))/colSums(((fit$sdev)^2)*t((Cargas^2)))
  perc_da_var_aj = as.data.frame(t(round(perc_da_var*100,2)))
  n_comp = length(perc_da_var_aj)
}

```

```

for( i in 1:(n_comp - 1)){
  perc_da_var_aj[,i +1] = perc_da_var_aj[,i] + perc_da_var_aj[,i + 1]
}
return(perc_da_var_aj)
}

#####
##### Funcao que ajusta a tabela de correlacao dos componentes principais
##### com as variaveis originais para impressao
##### fit objeto da classe princomp
#
#####

print_corr = function(fit){
  Cargas = loadings(fit)[]
  corr = (((fit$sdev)*t((Cargas)))/colSums(((fit$sdev)^2)*t((Cargas^2)))
  as.data.frame(corr)
  return(corr)
}

#####
##### Funcao que ajusta a tabela de resumo dos componentes principais
##### para impressao
##### fit objeto da classe princomp
#####

print_summary = function(fit){
  Summary_1 = rbind(fit$sdev^2, fit$sdev^2/sum(fit$sdev^2))
  n_comp = length(Summary_1[1,])
  Acum = array(-1,n_comp)
  Acum[1] = Summary_1[2,1]
  for( i in 1:(n_comp - 1)){
    Acum[i +1] = Acum[i] + Summary_1[2,i + 1]
  }
  Summary_1 = rbind(Summary_1,Acum)
  Summary_1 = as.data.frame(Summary_1)
  rownames(Summary_1) = c("Variancia","Proporcao da Variancia","Proporcao
    Acumulada da Variancia")
  return(Summary_1)
}

#####
##### Funcao que calcula a transformacao esferica para todas
##### as variaveis de uma base
##### Data base de dados com as variaveis a serem transformadas
#####

Tras_esferic = function(Data){
  Z = Data^(1/2)
  N_col = length(Z[,])
  N_lin = length(Z[,1])
}

```

```

W = matrix(0,N_lin ,N_col - 1)
Aux = 1
for ( i in 1:(N_col - 1)){
  W[ , i ] = acos(Z[ , i ]/Aux)
  Aux = Aux*sin(W[ , i ])
}
W = as.data.frame(W)
colnames(W) = colnames(Data)[-N_col]
return (W)
}

#####
##### funcões de Saídas Latex

#####
##### Função que imprime todas as tabelas da análises em formato Latex
##### FIT objeto da classe princomp
#####

Impri_Tabelas_comp_Princ = function(FIT){
  print("
  _____")
  print("Resumo")
  print("
  _____")
  Summary_1 = print.summary(FIT)
  Summary_impre <- xtable(Summary_1, digits = 2)
  print(Summary_impre, floating=TRUE)
  print("
  _____")
  print("Cargas")
  print("
  _____")
  Cargas = loadings(FIT)[]
  Summary_impre_1 <- xtable(Cargas, digits = 2)
  print(Summary_impre_1, floating=TRUE)
  print("
  _____")
  print("Percentual da varância da variável explicada pela componente")
  print("
  _____")
  comp_var = print_comp_var(FIT)
  Summary_impre_2 <- xtable(comp_var, digits = 2)
  print(Summary_impre_2, floating=TRUE)
  print("
  _____")
  print("Correlação Variável vs. Componente")
  print("
  _____")
  Corr_var = print_corr(FIT)
}

```

```

Summary_impre_3 <- xtable(Corr_var, digits = 2)
print(Summary_impre_3, floating=TRUE)
}

#####
##### Funcao de ajuste de componente Principais para todas as
Transformacoes
#####
## BASE Base com as variaveis para se realizar a analise
## VARS lista das variaveis que sera realizada a analise
## Trans lista das transformacoes que se deseja calcular sendo:
##      1 - Trad - Analise tradicional
##      2 - Exc - Exclusao de uma variavel
##      3 - Ref - Variavel de referencia
##      4 - Log - Transformacao Log
##      5 - Alr - Additive log-ratio transform
##      6 - Clr - Center log-ratio transform
##      7 - Esf - Transformacao Esferica
## ORDER_COMP Caso verdadeiro ordena as variaveis pela maior media
##             antes das analises
#####

Calc_princ_trnf = function(BASE,VARS,Trans = c(1:7),ORDER_COMP = TRUE){
  BASE_AUX = BASE[,VARS]
  N_comp = length(VARS)

  if (ORDER_COMP == TRUE) {
    Desc = as.data.frame(colSums(BASE_AUX)/length(BASE_AUX[,1]))
    names = rownames(Desc)[order(Desc[,1], decreasing = TRUE)]
    Data_Comp_Aj = BASE_AUX[,names]
  }
  FITS = vector("list",N_comp)
  #

  #### Trad - Analise tradicional
  #
  #

  if (1 %in% Trans) {
    data_Comp_trad = scale(Data_Comp_Aj)
    FITS[[1]] <- princomp(data_Comp_trad, cor=TRUE)
  }
  #

  #### Exc - Exclusao de uma variavel
  #
  #

  if (2 %in% Trans) {

```

```

FITS[[2]] <- princomp(Data_Comp_Aj[, -N_comp], cor=TRUE)
}
#
#####
##### Ref - Variavel de referencia
#
#####
if (3 %in% Trans) {
  data_Comp_ref = Data_Comp_Aj[, 1:(N_comp - 1)] / Data_Comp_Aj[, N_comp]
  FITS[[3]] <- princomp(data_Comp_ref, cor=TRUE)
}
#
#####
##### Log - Transformação Log
#
#####
if (4 %in% Trans) {
  data_Tranf_log = log(Data_Comp_Aj)
  FITS[[4]] <- princomp(data_Tranf_log, cor=TRUE)
}
#
#####
##### Alr - Additive log-ratio transform
#
#####
if (5 %in% Trans) {
  data_Tranf_alr = log(Data_Comp_Aj[, 1:(N_comp - 1)] / Data_Comp_Aj[, N_comp])
  FITS[[5]] <- princomp(data_Tranf_alr, cor=TRUE)
}
#
#####
##### Clr - Center log-ratio transform
#
#####
if (6 %in% Trans) {
  data_Tranf_clr = clr(Data_Comp_Aj)
  FITS[[6]] <- princomp(data_Tranf_clr, cor=TRUE)
}

```

```

#
#####
##### Esf - Transformação Esférica
#
#####

if (7 %in% Trans){
  data_Tranf_esfic = Tras_esferic(Data_Comp_Aj)
  FITS[[7]] <- princomp(data_Tranf_esfic, cor=TRUE)
}
names(FITS) = c("Trad", "Exc", "Ref", "Log", "Alr", "Clr", "Esf")
FITS = FITS[ !sapply(FITS, is.null) ]
return(FITS)
}

#####
##### Função calcula o numero necessário de componentes para explicar
##### X% da variância
#####
##### fit objeto da classe princomp
##### lim X% da variância explicada pela componente
#####

n_comp_var = function(fit, lim){
  aux = print_summary(fit)
  j = 1
  while(aux[3, j] < lim){
    j = j+1
  }
  return(j)
}

#####
##### Função tabula o % da variância explicado por cada componente
##### em cada cada técnica
#####
##### FITS lista contendo objetos da classe princomp
#####

tab_Var_Exp = function(FITS){
  N_FITS = length(FITS)
  N_var = length(print_summary(FITS[[1]])[1,])
  Saida = data.frame(array(-1,N_var))
  for(i in 1:N_FITS){
    aux = print_summary(FITS[[i]])
    N_t = length(aux)
    aux = array(c(as.numeric(aux[3,]), rep(0,N_var-N_t)))
    Saida[, i] = aux
  }
  rownames(Saida) = colnames(print_summary(FITS[[1]])[1,])
  colnames(Saida) = names(FITS)
  return(t(round(Saida*100,0)))
}

```

```

#####
##### Funcao tabula a quantidade de variaveis necessarias para explicar
##### % da variancia em cada tecnica
##### FITS lista contendo objetos da classe princomp
##### LIMITS lista dos percentuais que se deseja explicar
#####
tab_N_comp = function(FITS,LIMITS = c(0.6,0.7,0.8,0.9)){
  N_LIM = length(LIMITS)
  N_FITS = length(FITS)
  Saida = matrix(-1000,N_FITS,N_LIM)
  for(i in 1:N_FITS){
    for(j in 1:N_LIM){
      Saida[i,j] = n_comp_var(FITS[[i]],LIMITS[j])
    }
  }
  colnames(Saida) = paste(round(LIMITS*100),"%",sep = " ")
  rownames(Saida) = names(FITS)
  return(Saida)
}

#####
##### Funcao calcula o numero necessario de componentes para explicar
##### % da variancia em cada variavel
##### FITS lista contendo objetos da classe princomp
##### lim X% da variancia explicada pela componente
#####
n_comp_var_varialvel = function(fit,lim){
  aux = print_comp_var(fit)
  j = 1
  while(min(aux[,j]) < lim){
    j = j+1
  }
  return(j)
}

#####
##### Funcao calcula o numero necessario de componentes para explicar
##### % da variancia em cada variavel em cada tecnica
##### FITS lista contendo objetos da classe princomp
##### LIMITS lista dos percentuais que se deseja explicar
#####
tab_N_comp_variavel = function(FITS,LIMITS = c(60,70,80,90)){
  N_LIM = length(LIMITS)
  N_FITS = length(FITS)
  Saida = matrix(-1000,N_FITS,N_LIM)
  for(i in 1:N_FITS){
    for(j in 1:N_LIM){
      Saida[i,j] = n_comp_var_varialvel(FITS[[i]],LIMITS[j])
    }
  }
}

```

```

colnames(Saida) = paste(LIMITS,"%",sep = "")
rownames(Saida) = names(FITS)
return(Saida)
}

#####
##### Funcao tabula a ordem das variaveis com maiores cargas da
##### componente
#####
##### em cada tecnica
#####
##### FITS lista contendo objetos da classe princomp
#####
##### Comp Componente que se deseja tabular
#####

Lista_peso_var_comp = function(FITS,Comp){
  N_Tec = length(FITS)
  N_var = length(loadings(FITS[[1]])[,Comp])
  Var_Comp = data.frame(matrix(0,N_var,1))
  for(i in 1:N_Tec){
    aux = rownames(loadings(FITS[[i]])[]) [order(abs(t(loadings(FITS[[i]])
      [,Comp])),decreasing = TRUE)]
    n_aux = length(aux)
    aux = c(aux,rep("",N_var - n_aux))
    Var_Comp[,i] = aux
  }
  colnames(Var_Comp) = names(FITS)
  return(Var_Comp)
}
#
#####

#####
##### Funcao tabula e cria o grafico com a proporcao da carga da
##### variavel
#####
##### em relacao ao total da componente em cada tecnica
#####
##### FITS lista contendo objetos da classe princomp
#####
##### Comp Componente que se deseja tabular
#####
##### Comp Componente que se deseja tabular
#####
##### NOME String com o nome de saida do arquivo dos graficos
#####
##### pall paleta de cores da classe ggplot2 para se utilizar no
##### grafico
#
#####

Lista_peso_Carga_comp = function(FITS,Comp,NOME,pall){
  N_Tec = length(FITS)
  N_var = length(loadings(FITS[[1]])[,Comp])
  Var_Comp = data.frame(matrix(names(loadings(FITS[[1]])[,Comp]),N_var,1))
  colnames(Var_Comp) = c("VAR")
  for(i in 1:N_Tec){
    aux1= as.data.frame(names(loadings(FITS[[i]])[,Comp]))
    aux1[,2] = abs(loadings(FITS[[i]])[,Comp])
  }
}

```

```

colnames(aux1) = c("VAR", "V")
Var_Comp = merge(Var_Comp, aux1, by="VAR", all = TRUE, sort = FALSE)
colnames(Var_Comp)[i+1] = names(FITS)[i]
}
Var_Comp[is.na(Var_Comp)] <- 0
rownames(Var_Comp) = Var_Comp[, 1]
Var_Comp = Var_Comp[, -1]
Var_Comp = as.data.frame(t(Var_Comp))
Var_Comp_Pad = Var_Comp[, 1:N_var]/rowSums(Var_Comp[, 1:N_var])
Var_Comp_Pad$TEC = rownames(Var_Comp)

dat.aj <- melt(Var_Comp_Pad, id.vars = "TEC")
P = ggplot(dat.aj, aes(x = TEC, y = value, fill=variable)) + geom_bar(
  stat='identity')
P.labs = P + scale_fill_brewer(palette = pall) +
  labs(title = paste("Componente ", Comp, sep = ""), x="tecnicas", y =
    "Percentual do Peso das Cargas")

Title.format <- element_text(face="bold", size=24, hjust=0)
axis.format <- element_text(face="bold", size=16)
legend.format <- element_text(face="bold", size=16)
axis.format <- element_text(size=16)
p.fim = P.labs + theme(plot.title = Title.format, axis.title = axis.format,
  legend.title=element_blank(),
  legend.text = legend.format, axis.text = axis.format)

p.fim
ggsave(paste("Perc_carga_Comp_", Comp, "_", NOME, ".png", sep = ""), width =
  16, height = 7)
return(Var_Comp)
}

FITS = fits
#####
##### Funcao que imprime todas as tabelas da analises em formato Latex
##### FIT objeto da classe princomp
##### NOME String com o nome de saida das tabelas
#####
Impime_todas_as_Tabelas = function(FITS, NOME){
  N_var = length(FITS)
  for(i in 1:N_var){
    print("
      )
    print(paste(names(FITS)[i], "Tabelas de Analise formato latex"))
    print("
      )
    Impri_Tabelas_comp_Princ(FITS[[i]])
  }
}

```

```

}

print("

      ")
print("Tabela do percentual da variância explicada por cada componente
em cada técnica")
print("

      ")
t = t(tab_Var_Exp(FITS))
p = matrix(paste(t,"%",sep = ""),length(t[,1]),7)
colnames(p) = colnames(t)
rownames(p) = rownames(t)
impre_latex <- xtable(p, digits = 0)
print(impre_latex,floating=FALSE)
print("

      ")
print("Número de componentes necessárias para explicar X% da variância")
print("

      ")
impre_latex <- xtable(tab_N_comp(FITS), digits = 0)
print(impre_latex,floating=FALSE)

print("

      ")
print("Número de componentes necessárias para explicar X% da variância
de todas as variáveis")
print("

      ")
impre_latex <- xtable(tab_N_comp_variavel(FITS), digits = 0)
print(impre_latex,floating=FALSE)
print("

      ")
print("Lista das variáveis com maior peso para a componente 1")
print("

      ")
impre_latex <- xtable(Lista_peso_var_comp(FITS,1), digits = 0)
print(impre_latex,floating=FALSE)
print("

      ")
print("Lista das variáveis com maior peso para a componente 2")
print("

      ")
impre_latex <- xtable(Lista_peso_var_comp(FITS,1), digits = 0)
print(impre_latex,floating=FALSE)
}

#
#####
#####
```

```

##### Funcao gera todos os graficos da analise em todas as tecnicas
##### FITS lista contendo objetos da classe princomp
##### NOME String com o nome de saida do arquivo dos graficos
##### p all paleta de cores da classe ggplot2 para se utilizar no
##### grafico
#
#####
Gera_todos_Graf = function(FITS,NOME, p all) {
  N_var = length(FITS)
  for ( i in 1:N_var) {
    PLOT_Comp_Princ(paste("" ,names(FITS) [ i ] ,NOME, sep="_" ) ,FITS [[ i ]] ,FALSE)
  }
  Lista_peso_Carga_comp(FITS ,1 ,NOME, p all)
  Lista_peso_Carga_comp(FITS ,2 ,NOME, p all)
}
#
#####
##### Funcao Gera base de Dados Composicionais
##### baseados em p -1 uniformes e a p e pelo quanto falta para 1
##### N_obs: Numero de observacoes
##### N_Var: Nº de componentes
##### Medias: media dos p - varia eis
##### Desv: Desvio da media
#
#####
Comp_data_Sim_Unif = function(N_obs ,N_Var ,Medias ,Desv) {
  Aumento = 1.2
  N_gera = 0
  while (N_gera < N_obs) {
    Base = as.data.frame(matrix(-1,N_obs*Aumento,N_Var))
    for ( i in 1:(N_Var - 1)){
      Lim_inf = Medias[ i ] - sqrt(3*Desv[ i ])
      Lim_sup = Medias[ i ] + sqrt(3*Desv[ i ])
      Base[ ,i ] = r uniform(N_obs*Aumento,Lim_inf ,Lim_sup)
    }
    Base[ ,N_Var ] = 1 - rowSums(Base[ ,c(1:(N_Var -1))])
    N_gera = N_obs*Aumento - sum(Base[ ,N_Var]<0)
    Aumento = Aumento + 0.1
    if (Aumento > 2){
      return(print("Erro Mude os parametros da base"))
    }
  }
  Saida = Base[sample( which( Base[ ,N_Var] > 0 ) , N_obs ) ,]
  return(Saida)
}

```

```

#
#####
##### Funcao Gera base de Dados Composicionais
##### baseados em p -1 NOrmais e a p e pelo quanto falta para 1
##### N_obs: Numero de observacoes
##### N_Var: Nº de componentes
##### Medias: media dos p - variaveis
##### SD: Desvio padrao da Media
#
#####
Comp_data_Sim_Norm = function(N_obs,N_Var,Medias,SD){
  Aumento = 1.2
  N_gera = 0
  while (N_gera < N_obs){
    Base = as.data.frame(matrix(-1,N_obs*Aumento,N_Var))
    for( i in 1:(N_Var - 1)){
      Base[,i] = rtruncnorm(N_obs*Aumento, a=0, b=1, mean = Medias[i], sd = SD)
    }
    Base[,N_Var] = 1 - rowSums(Base[,c(1:(N_Var - 1))])
    N_gera = N_obs*Aumento - sum(Base[,N_Var]<0)
    Aumento = Aumento + 0.1
    if(Aumento > 3){
      return(print("Erro Mude os parametros da base"))
    }
  }
  Saida = Base[sample( which( Base[,N_Var] > 0 ), N_obs ),]
  return(Saida)
}

#
#####
##### Funcao Compara as posicoes das variaveis da componente i
##### baseados em p -1 NOrmais e a p e pelo quanto falta para 1
##### FITS lista contendo objetos da classe princomp
##### Comp Componente que se deseja tabular
#
#####

Compara_transf_pos = function(fits,Comp){
  tab = Lista_peso_var_comp(fits,Comp)
  N_Var = length(tab[,1])
  Sub = matrix(N_Var,7,7)
  colnames(Sub) = c("Trad","Exc","Ref","Log","Alr","Clr","Esf")
  rownames(Sub) = c("Trad","Exc","Ref","Log","Alr","Clr","Esf")
}

```

```

Sub[rownames(Sub) %in% c("Exc", "Ref", "Alr", "Esf"),] = N_Var - 1
Sub[, colnames(Sub) %in% c("Exc", "Ref", "Alr", "Esf")] = N_Var - 1

m_coin = matrix(-1, 7, 7)
colnames(m_coin) = colnames(tab)
rownames(m_coin) = colnames(tab)
for (i in 1:7) {
  for (j in i:7) {
    m_coin[i, j] = sum(tab[tab[, i] != "", & tab[, j] != "", i] != tab[tab[, i] != "",
      & tab[, j] != "", j])
    m_coin[j, i] = m_coin[i, j]
  }
}
m_coin[7, 7] = length(tab[tab[, 7] != "", 7])
return(m_coin)
}

#####
##### Funcao tabula a proporcao da carga da variavel
##### em relacao ao total da componente em cada tecnica
##### FITS lista contendo objetos da classe princomp
##### Comp Componente que se deseja tabular
#
#####

```

*##### Funcao tabula a proporcao da carga da variavel*

*##### em relacao ao total da componente em cada tecnica*

*##### FITS lista contendo objetos da classe princomp*

*##### Comp Componente que se deseja tabular*

*#####*

```

Lista_peso_Carga = function(FITS, Comp) {
  N_Tec = length(FITS)
  N_var = length(loadings(FITS[[1]])[, Comp])
  Var_Comp = data.frame(matrix(names(loadings(FITS[[1]])[, Comp]), N_var, 1))
  colnames(Var_Comp) = c("VAR")
  for(i in 1:N_Tec) {
    N_COMP = length(loadings(FITS[[i]])[1,])
    if(N_COMP<Comp) {
      aux1= as.data.frame(names(loadings(FITS[[i]])[, Comp - 1]))
      aux1[, 2] = 0
    } else {
      aux1= as.data.frame(names(loadings(FITS[[i]])[, Comp]))
      aux1[, 2] = loadings(FITS[[i]])[, Comp]
    }
    colnames(aux1) = c("VAR", "V")
    Var_Comp = merge(Var_Comp, aux1, by="VAR", all = TRUE, sort = FALSE)
    if(Var_Comp[1, i + 1]<0) {
      Var_Comp[, i + 1] = Var_Comp[, i + 1]*(-1)
    }
    colnames(Var_Comp)[i + 1] = names(FITS)[i]
  }
  rownames(Var_Comp) = Var_Comp[, 1]
}
```

```

Var_Comp = Var_Comp[, -1]
Var_Comp = as.data.frame(t(Var_Comp))
return(Var_Comp)
}

#
#####
##### Funcao Compara cria as tabelas para comprar as transformacoes
##### FITS lista contendo objetos da classe princomp
#
#####

Comparo_transf = function(fits){
  Var_Acum = tab_Var_Exp(fits)
  colnames(Var_Acum) = gsub("Comp.", "CP", colnames(Var_Acum))

  N_CP = length(colnames(Var_Acum))
  Div = matrix(N_CP, 7, 7)
  colnames(Div) = c("Trad", "Exc", "Ref", "Log", "Alr", "Clr", "Esf")
  rownames(Div) = c("Trad", "Exc", "Ref", "Log", "Alr", "Clr", "Esf")
  Div[rownames(Div) %in% c("Exc", "Ref", "Alr", "Esf"),] = N_CP - 1
  Div[, colnames(Div) %in% c("Exc", "Ref", "Alr", "Esf")] = N_CP - 1

  for (i in 1:N_CP){
    if (i == 1){
      Carga_Var = as.matrix(vegdist(Lista_peso_Carga(fits, i), method =
        "manhattan", na.rm = TRUE, diag = FALSE)) / Div
    } else{
      Carga_Var = Carga_Var + as.matrix(vegdist(Lista_peso_Carga(fits, i),
        method = "manhattan", na.rm = TRUE, diag = FALSE)) / Div
    }
  }

  Carga_Var = Carga_Var / Div
  Dist_carga = as.dist(Carga_Var)

  return(list(Var_Acum = Var_Acum, Dist_carga = Dist_carga))
}

#
#####
##### Funcao simula bases de uma dirichlet e
##### compara os resultados das transformacoes
##### N_Bases: numero de bases para gerar
##### N_obs: numero de bases para gerar

```

```

##### N_Var: numero de bases para gerar
##### Conc_Param: numero de bases para gerar
#
#####
N_bases_Comp_data_Sim_D = function(N_Bases ,N_obs ,N_Var ,Conc_Param) {
  Saidas = list ()

  Div = matrix(N_Var ,7 ,7)
  colnames(Div) = c("Trad" , "Exc" , "Ref" , "Log" , "Alr" , "Clr" , "Esf")
  rownames(Div) = c("Trad" , "Exc" , "Ref" , "Log" , "Alr" , "Clr" , "Esf")
  Div[rownames(Div) %in% c("Exc" , "Ref" , "Alr" , "Esf") ,] = N_Var - 1
  Div[ , colnames(Div) %in% c("Exc" , "Ref" , "Alr" , "Esf")] = N_Var - 1

  timestamp()
  for ( i in 1:N_Bases ) {
    base = as.data.frame( rdirichlet(N_obs ,Conc_Param) )
    fits = Calc_princ_tranf(base ,c(1:N_Var))
    if ( i == 1) {
      Aux = Compara_transf(fits)
      ##### Variancia Acumulada
      Saidas$Var_Acum = as.matrix(Aux$Var_Acum)
      Saidas$Var_Acum_Quad = Saidas$Var_Acum^2
      ##### Diferenca das Cargas
      Saidas$Dist_carga = as.matrix(Aux$Dist_carga)/Div
      Saidas$Dist_carga_Quad = Saidas$Dist_carga^2
    } else {
      Aux = Compara_transf(fits)
      ##### Variancia Acumulada
      Saidas$Var_Acum = Saidas$Var_Acum + as.matrix(Aux$Var_Acum)
      Saidas$Var_Acum_Quad = Saidas$Var_Acum_Quad + as.matrix(Aux$Var_Acum)^2
      ##### Diferenca das Cargas
      Saidas$Dist_carga = Saidas$Dist_carga + (as.matrix(Aux$Dist_carga)/
        Div)
      Saidas$Dist_carga_Quad = Saidas$Dist_carga_Quad + (as.matrix(Aux$Dist_carga)/
        Div)^2
    }
    if ( i %% 100==0) {
      print(i)
    }
  }
  ##### Variancia Acumulada
  Saidas$Var_Acum_Med = Saidas$Var_Acum/N_Bases
  Saidas$Var_Acum_Var = (Saidas$Var_Acum_Quad - (Saidas$Var_Acum^2)/N_Bases)/
    (N_Bases)
  Saidas$Var_Acum_sd = sqrt(Saidas$Var_Acum_Var)
}

```

```

Saidas$Var_Acum_Lim_Inf = Saidas$Var_Acum_Med - 1.96*(Saidas$Var_Acum_sd
  /sqrt(N_Bases))
Saidas$Var_Acum_Lim_Sup = Saidas$Var_Acum_Med + 1.96*(Saidas$Var_Acum_sd
  /sqrt(N_Bases))
#####Diferenca das Cargas
Saidas$Dist_carga_Med = Saidas$Dist_carga/N_Bases
Saidas$Dist_carga_Var = (Saidas$Dist_carga_Quad - (Saidas$Dist_carga^2)/
  N_Bases)/(N_Bases)
Saidas$Dist_carga_sd = sqrt(Saidas$Dist_carga_Var)
Saidas$Dist_carga_Lim_Inf = Saidas$Dist_carga_Med - 1.96*(Saidas$Dist_carga_sd/sqrt(N_Bases))
Saidas$Dist_carga_Lim_Sup = Saidas$Dist_carga_Med + 1.96*(Saidas$Dist_carga_sd/sqrt(N_Bases))
timestamp()
return(Saidas)
}

#
#####
##### Funcao simula bases de uma uniforme e
##### compara os resultados das transformacoes
##### N_Bases: numero de bases para gerar
##### N_obs: numero de bases para gerar
##### N_Var: numero de bases para gerar
##### Medias: numero de bases para gerar
##### Desv: desvios da media
#
#####
N_bases_Comp_data_Sim_U = function(N_Bases,N_obs,N_Var,Medias,Desv){
  Saidas = list()

  Div = matrix(N_Var,7,7)
  colnames(Div) = c("Trad","Exc","Ref","Log","Alr","Clr","Esf")
  rownames(Div) = c("Trad","Exc","Ref","Log","Alr","Clr","Esf")
  Div[rownames(Div) %in% c("Exc","Ref","Alr","Esf"),] = N_Var - 1
  Div[,colnames(Div) %in% c("Exc","Ref","Alr","Esf")] = N_Var - 1

  timestamp()
  for ( i in 1:N_Bases){
    base = Comp_data_Sim_Unif(N_obs,N_Var,Medias,Desv)
    fits = Calc_princ_trnf(base,c(1:N_Var))
    if (i == 1){
      Aux = Compara_transf(fits)
      ####Variancia Acumulada
      Saidas$Var_Acum = as.matrix(Aux$Var_Acum)
      Saidas$Var_Acum_Quad = Saidas$Var_Acum^2
      #####Diferenca das Cargas
    }
  }
}

```

```

Saidas$Dist_carga = as.matrix(Aux$Dist_carga)/Div
Saidas$Dist_carga_Quad = Saidas$Dist_carga^2
} else {
  Aux = Compara_transf(fits)
  #####Variancia Acumulada
  Saidas$Var_Acum = Saidas$Var_Acum + as.matrix(Aux$Var_Acum)
  Saidas$Var_Acum_Quad = Saidas$Var_Acum_Quad + as.matrix(Aux$Var_Acum)^2
  #####Diferenca das Cargas
  Saidas$Dist_carga = Saidas$Dist_carga + (as.matrix(Aux$Dist_carga) /
    Div)
  Saidas$Dist_carga_Quad = Saidas$Dist_carga_Quad + (as.matrix(Aux$Dist_carga)/Div)^2
}
if (i %% 100==0){
  print(i)
}
#####
#####Variancia Acumulada
Saidas$Var_Acum_Med = Saidas$Var_Acum/N_Bases
Saidas$Var_Acum_Var = (Saidas$Var_Acum_Quad - (Saidas$Var_Acum^2)/N_Bases) / (N_Bases)
Saidas$Var_Acum_sd = sqrt(Saidas$Var_Acum_Var)
Saidas$Var_Acum_Lim_Inf = Saidas$Var_Acum_Med - 1.96*(Saidas$Var_Acum_sd /
  sqrt(N_Bases))
Saidas$Var_Acum_Lim_Sup = Saidas$Var_Acum_Med + 1.96*(Saidas$Var_Acum_sd /
  sqrt(N_Bases))
#####Diferenca das Cargas
Saidas$Dist_carga_Med = Saidas$Dist_carga/N_Bases
Saidas$Dist_carga_Var = (Saidas$Dist_carga_Quad - (Saidas$Dist_carga^2)/
  N_Bases) / (N_Bases)
Saidas$Dist_carga_sd = sqrt(Saidas$Dist_carga_Var)
Saidas$Dist_carga_Lim_Inf = Saidas$Dist_carga_Med - 1.96*(Saidas$Dist_carga_sd /
  sqrt(N_Bases))
Saidas$Dist_carga_Lim_Sup = Saidas$Dist_carga_Med + 1.96*(Saidas$Dist_carga_sd /
  sqrt(N_Bases))
timestamp()
return(Saidas)
}
#
#####
##### Funcao simula bases de uma normais e
##### compara os resultados das transformacoes
##### N_Bases: numero de bases para gerar
##### N_obs: numero de bases para gerar
##### N_Var: numero de bases para gerar
##### Medias: numero de bases para gerar

```

```

##### Desv: desvios da media
#
#####
N_bases_Comp_data_Sim_N = function(N_Bases ,N_obs ,N_Var ,Medias ,Desv){
  Saidas = list ()

  Div = matrix(N_Var ,7 ,7)
  colnames(Div) = c("Trad" , "Exc" , "Ref" , "Log" , "Alr" , "Clr" , "Esf")
  rownames(Div) = c("Trad" , "Exc" , "Ref" , "Log" , "Alr" , "Clr" , "Esf")
  Div[rownames(Div) %in% c("Exc" , "Ref" , "Alr" , "Esf") ,] = N_Var - 1
  Div[ ,colnames(Div) %in% c("Exc" , "Ref" , "Alr" , "Esf")] = N_Var - 1

  timestamp()
  for ( i in 1:N_Bases){
    base = Comp_data_Sim_Norm(N_obs ,N_Var ,Medias ,Desv)
    fits = Calc_princ_trnf(base ,c(1:N_Var))
    if ( i == 1){
      Aux = Compara_transf(fits)
      ##### Variância Acumulada
      Saidas$Var_Acum = as.matrix(Aux$Var_Acum)
      Saidas$Var_Acum_Quad = Saidas$Var_Acum^2
      ##### Diferença das Cargas
      Saidas$Dist_carga = as.matrix(Aux$Dist_carga)/Div
      Saidas$Dist_carga_Quad = Saidas$Dist_carga^2
    } else{
      Aux = Compara_transf(fits)
      ##### Variância Acumulada
      Saidas$Var_Acum = Saidas$Var_Acum + as.matrix(Aux$Var_Acum)
      Saidas$Var_Acum_Quad = Saidas$Var_Acum_Quad + as.matrix(Aux$Var_Acum)^2
      ##### Diferença das Cargas
      Saidas$Dist_carga = Saidas$Dist_carga + (as.matrix(Aux$Dist_carga)/
        Div)
      Saidas$Dist_carga_Quad = Saidas$Dist_carga_Quad + (as.matrix(Aux$Dist_carga)/
        Div)^2
    }
    if ( i %% 100 == 0){
      print(i)
    }
  }
  ##### Variância Acumulada
  Saidas$Var_Acum_Med = Saidas$Var_Acum/N_Bases
  Saidas$Var_Acum_Var = (Saidas$Var_Acum_Quad - (Saidas$Var_Acum^2)/N_Bases)/(N_Bases)
  Saidas$Var_Acum_sd = sqrt(Saidas$Var_Acum_Var)
  Saidas$Var_Acum_Lim_Inf = Saidas$Var_Acum_Med - 1.96*(Saidas$Var_Acum_sd/
    sqrt(N_Bases))
}

```

```

Saidas$Var_Acum_Lim_Sup = Saidas$Var_Acum_Med + 1.96*(Saidas$Var_Acum_sd
  /sqrt(N_Bases))
#####Diferenca das Cargas
Saidas$Dist_carga_Med = Saidas$Dist_carga/N_Bases
Saidas$Dist_carga_Var = (Saidas$Dist_carga_Quad - (Saidas$Dist_carga^2)/
  N_Bases)/(N_Bases)
Saidas$Dist_carga_sd = sqrt(Saidas$Dist_carga_Var)
Saidas$Dist_carga_Lim_Inf = Saidas$Dist_carga_Med - 1.96*(Saidas$Dist_carga_sd/sqrt(N_Bases))
Saidas$Dist_carga_Lim_Sup = Saidas$Dist_carga_Med + 1.96*(Saidas$Dist_carga_sd/sqrt(N_Bases))
timestamp()
return(Saidas)
}

imprime_Saida_comp = function(Sim,NOME) {
  #####Intervalo da Variancia acumulada
  AUX_INF = t(Sim$Var_Acum_Lim_Inf)
  AUX_SUP = t(Sim$Var_Acum_Lim_Sup)
  N_comp = length(AUX_SUP[,1])
  print_var = as.data.frame(matrix(gsub("\\" , " ,",
    paste("[ ",round(AUX_INF,2) , " ; " ,
    round(AUX_SUP,2) ,
    "] " ,sep = " ") ) ,N_comp ,7))
  colnames(print_var) = colnames(AUX_SUP)
  rownames(print_var) = rownames(AUX_SUP)
  print("-----")
  print("Intervalo de confiança para a variancia explicada")
  print("-----")
  impre_inter_var<- xtable(print_var,label = paste("tab:Sim_InterVar_",
    NOME,sep = " ") ,
    align = c("l",rep("c",7)))
  print(impre_inter_var)

  print("-----")
  print("Matriz de medias de diferenças")
  print("-----")
  impre_Dif_Var<- xtable(Sim$Dist_carga_Med,label = paste("tab:Sim_Matriz_",
    " ,NOME,sep = " ") ,
    align = c("l",rep("c",7)), digits = 4)
  print(impre_Dif_Var)

t1.1 = hclust(as.dist(Sim$Dist_carga_Med))
png(paste("Dend_Carg" ,NOME ,".png" ,sep = " ") ,width = 600, height = 450)
par(mar=c(0, 4, 4, 2)) # c(bottom, left, top, right)
plot(t1.1,xlab = "",main = "",ylab = "Dif .")
dev.off()
}

```

## Apêndice C

### Bases de Dados Utilizadas

## C.1 Base 1

A base de dados foi extraída do artigo original Kuhn *et al.* (2000), os elementos Cu, Ni e Zn não foram utilizados e os dados foram padronizados para que a soma dos elementos restantes somassem 1. A base utilizada na análise se encontra nas Tabelas C.1 a C.3.

**Tabela C.1:** *Base de Dados 1*

ID	Si%	Ti%	Al%	Fe%	Mg%	K%	Mn%
54GK/0	0,4665	0,0099	0,1364	0,2419	0,1237	0,0085	0,0131
	10	0,4534	0,0088	0,1278	0,2622	0,1202	0,0120
	20	0,4948	0,0102	0,1451	0,2171	0,1115	0,0113
	30	0,4468	0,0088	0,1306	0,2680	0,1180	0,0141
	40	0,4533	0,0087	0,1326	0,2650	0,1142	0,0124
	49	0,4478	0,0085	0,1288	0,2797	0,1083	0,0111
72PC/0	0,2524	0,0062	0,1264	0,4260	0,1421	0,0124	0,0344
	10	0,2010	0,0084	0,1609	0,4088	0,1668	0,0198
	20	0,2159	0,0082	0,1532	0,4268	0,1446	0,0147
	30	0,2497	0,0085	0,1551	0,3803	0,1529	0,0187
	40	0,1923	0,0076	0,1596	0,4068	0,1818	0,0181
	50	0,2209	0,0081	0,1635	0,3965	0,1667	0,0121
	60	0,1686	0,0110	0,1621	0,4166	0,1796	0,0243
	70	0,2655	0,0088	0,1693	0,3694	0,1407	0,0148
	80	0,2380	0,0102	0,1740	0,3696	0,1594	0,0187
	90	0,2782	0,0081	0,1627	0,3711	0,1411	0,0116
	100	0,2402	0,0102	0,1493	0,4083	0,1529	0,0105
	110	0,2479	0,0077	0,1574	0,3817	0,1564	0,0185
	120	0,2613	0,0095	0,1519	0,4011	0,1221	0,0242
	130	0,2462	0,0086	0,1294	0,4360	0,1259	0,0233
	140	0,2812	0,0063	0,1461	0,3649	0,1444	0,0305
81PC/0	150	0,2928	0,0127	0,1562	0,3544	0,1358	0,0231
	160	0,2896	0,0074	0,1380	0,3966	0,1201	0,0218
	170	0,2695	0,0079	0,1303	0,4276	0,1211	0,0153
	180	0,2906	0,0086	0,1285	0,4089	0,1241	0,0121
	200	0,3384	0,0062	0,1281	0,3809	0,1098	0,0113
	210	0,3413	0,0079	0,1249	0,3785	0,1081	0,0115
	220	0,2529	0,0100	0,1470	0,4038	0,1303	0,0122
	230	0,2611	0,0065	0,1547	0,3877	0,1312	0,0183
	240	0,2394	0,0082	0,1572	0,3951	0,1317	0,0250
	250	0,3337	0,0090	0,1708	0,3212	0,1089	0,0223
	260	0,2438	0,0100	0,1677	0,3838	0,1235	0,0288
	270	0,2360	0,0116	0,1621	0,3985	0,1385	0,0095
	280	0,2715	0,0113	0,1749	0,3589	0,1171	0,0221
	290	0,2750	0,0119	0,1636	0,3640	0,1159	0,0242
	300	0,2044	0,0096	0,1587	0,4398	0,1177	0,0166
	310	0,2097	0,0114	0,1550	0,4303	0,1165	0,0247
	320	0,2271	0,0099	0,1561	0,4123	0,1232	0,0193
	330	0,2876	0,0079	0,1764	0,3438	0,1306	0,0172
	340	0,3019	0,0097	0,1764	0,3380	0,1210	0,0171
	350	0,2998	0,0106	0,1578	0,3586	0,1117	0,0247
	360	0,2852	0,0089	0,1607	0,3797	0,1067	0,0197
	370	0,2168	0,0126	0,1608	0,4292	0,1226	0,0138
	380	0,2042	0,0103	0,1532	0,4487	0,1255	0,0168
	390	0,2039	0,0104	0,1545	0,4398	0,1271	0,0183
	400	0,1999	0,0095	0,1524	0,4488	0,1243	0,0163
	410	0,2829	0,0132	0,1859	0,3440	0,1226	0,0161
	420	0,2082	0,0122	0,1804	0,4069	0,1344	0,0189
	430	0,1314	0,0098	0,1779	0,4476	0,1564	0,0274
	440	0,1749	0,0114	0,1652	0,4492	0,1405	0,0146
	450	0,1415	0,0078	0,1516	0,4916	0,1472	0,0176

**Tabela C.2:** Continuação da Base de Dados 1

ID	Si%	Ti%	Al%	Fe%	Mg%	K%	Mn%
240	0,1477	0,0113	0,1517	0,4859	0,1407	0,0202	0,0425
250	0,2383	0,0106	0,1643	0,4076	0,1357	0,0106	0,0327
260	0,3173	0,0088	0,1664	0,3484	0,1108	0,0208	0,0275
270	0,3379	0,0082	0,1666	0,3376	0,0996	0,0214	0,0285
280	0,3302	0,0101	0,1518	0,3665	0,0924	0,0196	0,0294
290	0,3208	0,0077	0,1394	0,3961	0,0886	0,0164	0,0308
300	0,2917	0,0092	0,1475	0,3978	0,1055	0,0200	0,0281
310	0,3304	0,0094	0,1759	0,3290	0,1071	0,0257	0,0226
320	0,3795	0,0078	0,1818	0,2723	0,1112	0,0278	0,0196
340	0,3862	0,0085	0,1644	0,3045	0,0981	0,0191	0,0191
350	0,3978	0,0077	0,1512	0,3237	0,0845	0,0162	0,0188
360	0,3936	0,0088	0,1419	0,3320	0,0818	0,0200	0,0217
370	0,4652	0,0133	0,1635	0,2399	0,0921	0,0138	0,0122
380	0,4703	0,0122	0,1649	0,2357	0,0913	0,0124	0,0132
390	0,4675	0,0120	0,1581	0,2446	0,0910	0,0140	0,0130
400	0,4447	0,0105	0,1523	0,2804	0,0826	0,0152	0,0144
410	0,4385	0,0112	0,1524	0,2815	0,0834	0,0180	0,0150
95PC/0	0,2803	0,0072	0,1018	0,4504	0,1074	0,0141	0,0388
10	0,2989	0,0075	0,1097	0,4138	0,1250	0,0135	0,0315
20	0,3105	0,0066	0,1130	0,3716	0,1482	0,0180	0,0321
30	0,3414	0,0087	0,1304	0,3596	0,1104	0,0205	0,0290
40	0,3245	0,0086	0,1289	0,3761	0,1228	0,0089	0,0302
50	0,2909	0,0097	0,1244	0,3923	0,1326	0,0157	0,0344
60	0,2584	0,0104	0,1377	0,4046	0,1377	0,0112	0,0398
70	0,2622	0,0104	0,1377	0,4009	0,1393	0,0120	0,0375
80	0,2138	0,0061	0,1309	0,4588	0,1360	0,0135	0,0410
90	0,1683	0,0108	0,1226	0,4697	0,1667	0,0204	0,0416
100	0,2133	0,0055	0,1287	0,4121	0,1793	0,0242	0,0370
110	0,2426	0,0072	0,1239	0,4206	0,1517	0,0182	0,0359
120	0,2204	0,0088	0,1127	0,4631	0,1400	0,0157	0,0392
130	0,2839	0,0053	0,1211	0,4244	0,1193	0,0109	0,0351
150	0,2827	0,0063	0,1217	0,4452	0,1044	0,0084	0,0313
160	0,2363	0,0047	0,0975	0,4936	0,1111	0,0165	0,0402
170	0,3063	0,0063	0,1109	0,4416	0,0956	0,0090	0,0303
180	0,3421	0,0066	0,1073	0,3978	0,0971	0,0194	0,0298
190	0,3374	0,0067	0,1136	0,3954	0,1041	0,0140	0,0288
200	0,2951	0,0075	0,1071	0,4457	0,1024	0,0099	0,0323
210	0,3449	0,0067	0,1053	0,4116	0,0988	0,0045	0,0281
220	0,3103	0,0070	0,1048	0,4357	0,1058	0,0070	0,0294
230	0,3441	0,0078	0,1179	0,4013	0,0933	0,0086	0,0271
240	0,2917	0,0070	0,1123	0,4410	0,0981	0,0193	0,0305
250	0,2633	0,0068	0,0993	0,4645	0,1150	0,0191	0,0320
260	0,2770	0,0070	0,1070	0,4666	0,1073	0,0050	0,0301
270	0,2887	0,0087	0,1093	0,4378	0,1109	0,0133	0,0314
280	0,2344	0,0077	0,1154	0,4729	0,1217	0,0161	0,0319
290	0,1948	0,0075	0,1105	0,4897	0,1386	0,0231	0,0359
300	0,1994	0,0053	0,1041	0,5054	0,1272	0,0208	0,0379
310	0,2099	0,0096	0,1135	0,4811	0,1342	0,0133	0,0384
320	0,1942	0,0049	0,1174	0,4804	0,1495	0,0146	0,0390
330	0,1360	0,0064	0,1193	0,5311	0,1459	0,0172	0,0442
340	0,1980	0,0084	0,1161	0,4925	0,1324	0,0137	0,0390
350	0,1507	0,0123	0,1225	0,4767	0,1666	0,0220	0,0492
360	0,1500	0,0098	0,1237	0,4804	0,1711	0,0191	0,0459
370	0,1634	0,0069	0,1257	0,5135	0,1378	0,0112	0,0414
380	0,1768	0,0071	0,1137	0,5223	0,1248	0,0161	0,0391
390	0,2287	0,0087	0,1198	0,4549	0,1327	0,0161	0,0392
400	0,2152	0,0073	0,1194	0,4756	0,1228	0,0209	0,0388

**Tabela C.3:** Continuação da Base de Dados 1

ID	Si%	Ti%	Al%	Fe%	Mg%	K%	Mn%
410	0,2640	0,0077	0,1151	0,4564	0,1146	0,0077	0,0345
420	0,2752	0,0074	0,1240	0,4420	0,1095	0,0121	0,0298
430	0,2735	0,0098	0,1327	0,4313	0,1083	0,0173	0,0271
439	0,3353	0,0080	0,1374	0,3794	0,1007	0,0162	0,0231
449	0,3914	0,0093	0,1288	0,3332	0,0968	0,0193	0,0212
459	0,4739	0,0087	0,1215	0,2827	0,0861	0,0110	0,0161
470	0,4984	0,0088	0,1268	0,2452	0,0876	0,0194	0,0138
480	0,5047	0,0093	0,1310	0,2271	0,0930	0,0208	0,0141
490	0,5441	0,0086	0,1337	0,1836	0,0953	0,0226	0,0122
500	0,5365	0,0075	0,1203	0,1825	0,1272	0,0172	0,0088
510	0,5428	0,0078	0,1143	0,1872	0,1226	0,0165	0,0089
520	0,5619	0,0091	0,1270	0,1754	0,0980	0,0199	0,0087
525	0,5703	0,0096	0,1287	0,1703	0,0881	0,0252	0,0077
530	0,5785	0,0104	0,1257	0,1697	0,0826	0,0260	0,0070
535	0,5711	0,0087	0,1168	0,1754	0,0976	0,0227	0,0078
540	0,5379	0,0084	0,1146	0,1973	0,1111	0,0213	0,0094
550	0,5433	0,0092	0,1173	0,2056	0,0932	0,0212	0,0101
560	0,5965	0,0123	0,1208	0,1753	0,0621	0,0255	0,0075
570	0,5687	0,0102	0,1296	0,1834	0,0722	0,0283	0,0076
580	0,5527	0,0086	0,1235	0,2001	0,0844	0,0219	0,0088
26PC/0	0,2968	0,0078	0,0940	0,4155	0,1095	0,0276	0,0488
20	0,2968	0,0065	0,1161	0,3991	0,1118	0,0254	0,0443
30	0,3324	0,0073	0,1078	0,3775	0,1084	0,0265	0,0401
40	0,3436	0,0084	0,1121	0,3470	0,1216	0,0342	0,0331
50	0,3330	0,0081	0,1144	0,3563	0,1164	0,0364	0,0354
60	0,3057	0,0099	0,1106	0,4058	0,1013	0,0235	0,0432
75	0,3182	0,0085	0,1170	0,3655	0,1183	0,0316	0,0408
85	0,3524	0,0084	0,1265	0,3241	0,1181	0,0319	0,0386
95	0,3263	0,0099	0,1201	0,3480	0,1269	0,0316	0,0372
105	0,3330	0,0089	0,1333	0,3387	0,1255	0,0282	0,0324
115	0,3329	0,0089	0,1240	0,3380	0,1355	0,0285	0,0323
27PC/0	0,4209	0,0086	0,1269	0,2817	0,1129	0,0238	0,0251
10	0,4211	0,0088	0,1243	0,2872	0,1087	0,0244	0,0256
20	0,4321	0,0084	0,1369	0,2704	0,0993	0,0285	0,0243
30	0,4225	0,0085	0,1455	0,2675	0,1036	0,0278	0,0245
40	0,4246	0,0083	0,1403	0,2774	0,0987	0,0253	0,0253
50	0,4100	0,0091	0,1371	0,2881	0,1047	0,0267	0,0243
60	0,4193	0,0082	0,1378	0,2808	0,1030	0,0277	0,0232
70	0,4512	0,0093	0,1523	0,2402	0,1014	0,0274	0,0181
80	0,4574	0,0093	0,1562	0,2257	0,1098	0,0259	0,0157
90	0,4455	0,0094	0,1500	0,2443	0,1082	0,0241	0,0185
100	0,4537	0,0097	0,1533	0,2301	0,1102	0,0259	0,0172
110	0,4588	0,0103	0,1477	0,2321	0,1070	0,0257	0,0183
120	0,4879	0,0101	0,1652	0,2033	0,0976	0,0218	0,0140
130	0,4859	0,0110	0,1627	0,2031	0,1066	0,0179	0,0129
140	0,5129	0,0116	0,1764	0,1734	0,1069	0,0108	0,0081
150	0,4937	0,0109	0,1656	0,1947	0,1087	0,0157	0,0106
160	0,4970	0,0109	0,1705	0,1895	0,1111	0,0118	0,0092
170	0,4520	0,0093	0,1471	0,2419	0,1137	0,0205	0,0156
180	0,4063	0,0102	0,1291	0,3117	0,0924	0,0270	0,0234
185	0,4461	0,0096	0,1426	0,2599	0,0981	0,0252	0,0185
200	0,4910	0,0136	0,1636	0,1920	0,1197	0,0114	0,0088
210	0,4178	0,0095	0,1326	0,2934	0,0968	0,0256	0,0243

## C.2 Base 2

A base de dados foi extraída do artigo original Grantham e Velbel (1988) e se encontra nas Tabelas C.4 e C.5.

**Tabela C.4: Base de Dados 2**

Amostra	tipo de grão	Qm%	Qp%	Rf%	M%
2	c	0,1340	0,4000	0,4370	0,0280
2	c	0,1250	0,4000	0,4410	0,0340
2	c	0,1410	0,3840	0,4300	0,0460
2	m	0,3530	0,2570	0,2770	0,1150
2	m	0,3890	0,1790	0,3100	0,1230
2	m	0,3640	0,2000	0,3340	0,1030
2	f	0,5230	0,0600	0,2520	0,1640
2	f	0,5000	0,0520	0,2600	0,1880
2	f	0,4840	0,0590	0,2800	0,1770
10	c	0,1720	0,3700	0,3960	0,0620
10	c	0,2080	0,3510	0,3510	0,0880
10	c	0,1650	0,4030	0,3450	0,0860
10	m	0,4070	0,2640	0,1770	0,1510
10	m	0,3810	0,2640	0,2220	0,1340
10	m	0,4000	0,2450	0,1500	0,2070
10	f	0,6270	0,0550	0,1180	0,2000
10	f	0,6440	0,0790	0,1130	0,1620
10	f	0,5500	0,0610	0,1110	0,2790
34	c	0,1870	0,2920	0,3530	0,1680
34	c	0,1530	0,3360	0,3860	0,1230
34	c	0,1990	0,3360	0,3510	0,1130
34	m	0,4070	0,1660	0,1740	0,2520
34	m	0,4250	0,2030	0,1930	0,1770
34	m	0,3960	0,1770	0,1770	0,2470
34	f	0,6460	0,0390	0,1080	0,2060
34	f	0,6830	0,0580	0,1100	0,1470
34	f	0,6220	0,0440	0,1170	0,2170
41	c	0,1970	0,2590	0,4130	0,1300
41	c	0,2880	0,2080	0,3670	0,1360
41	c	0,1880	0,2850	0,3950	0,1310
41	m	0,4520	0,0960	0,2730	0,1790
41	m	0,4500	0,1370	0,2480	0,1650
41	m	0,4740	0,0960	0,2710	0,1600
41	f	0,5290	0,0470	0,1980	0,2250
41	f	0,4830	0,0470	0,1880	0,2810
41	f	0,5320	0,0740	0,1830	0,2100
13	c	0,3360	0,2120	0,4450	0,0070
13	c	0,2740	0,2290	0,4900	0,0070
13	c	0,2910	0,2420	0,4580	0,0090
13	m	0,5740	0,1090	0,1990	0,1170
13	m	0,5930	0,1010	0,2140	0,0930
13	m	0,5550	0,0830	0,2110	0,1510
13	f	0,6300	0,0470	0,1250	0,1980
13	f	0,6470	0,0320	0,1190	0,2020
13	f	0,6440	0,0430	0,1070	0,2050
27	c	0,2460	0,3450	0,3670	0,0420
27	c	0,2030	0,2800	0,3920	0,1250
27	c	0,1870	0,3910	0,3520	0,0700
27	m	0,5070	0,1030	0,1760	0,2130
27	m	0,4740	0,1520	0,1690	0,2040
27	m	0,5130	0,1050	0,1750	0,2060
27	f	0,6500	0,0270	0,0860	0,2370
27	f	0,6600	0,0170	0,0790	0,2440
27	f	0,6480	0,0280	0,0810	0,2430

**Tabela C.5:** Continuação da Base de Dados 2

Amostra	tipo de grão	Qm%	Qp%	Rf%	M%
32	c	0,4010	0,3090	0,2600	0,0310
32	c	0,3950	0,3070	0,2650	0,0330
32	c	0,4220	0,2700	0,2360	0,0710
32	m	0,5600	0,0930	0,1660	0,1810
32	m	0,5840	0,0940	0,1690	0,1540
32	m	0,6240	0,1150	0,1340	0,1200
32	f	0,6750	0,0260	0,0630	0,2350
32	f	0,6620	0,0360	0,0650	0,2370
32	f	0,6620	0,0370	0,0630	0,2390
37	c	0,1710	0,2320	0,5440	0,0520
37	c	0,1400	0,1890	0,5770	0,0940
37	c	0,1370	0,2860	0,5300	0,0470
37	m	0,4510	0,0900	0,2640	0,1940
37	m	0,4400	0,1340	0,2800	0,1450
37	m	0,4070	0,1290	0,2710	0,1930
37	f	0,5540	0,0400	0,1220	0,2840
37	f	0,5470	0,0590	0,1280	0,2660
37	f	0,5820	0,0410	0,1340	0,2420

### C.3 Base 3

A base de dados foi coletada em Tolosana-Delgado e von Eynatten (2010) e a base analítica foi extraída de Van den Boogaart e Tolosana-Delgado (2013) e Gentleman *et al.* (2013) e encontra-se nas Tabelas C.6 e C.7.. Ela é composta pelos seguintes óxidos:  $SiO_2$ ,  $TiO_2$ ,  $Al_2O_3$ ,  $MnO$ ,  $MgO$ ,  $CaO$ ,  $Na_2O$ ,  $K_2O$ ,  $P_2O_5$ ,  $Fe_2O_3$  e outras substâncias que não foram capazes de serem medidas.

**Tabela C.6:** Base de Dados 3

Amostra	Tipo do Grão	$SiO_2$ %	$TiO_2$ %	$Al_2O_3$ %	$MnO$ %	$MgO$ %	$CaO$ %	$Na_2O$ %	$K_2O$	$P_2O_5$ %	$Fe_2O_3$ %	Outros
RT1-1	-1	0,7340	0,0022	0,1419	0,0004	0,0048	0,0104	0,0433	0,0418	0,0006	0,0174	0,0032
RT1-1	0	0,7386	0,0026	0,1400	0,0004	0,0051	0,0114	0,0432	0,0395	0,0007	0,0172	0,0013
RT1-1	1	0,7385	0,0030	0,1385	0,0004	0,0055	0,0115	0,0431	0,0382	0,0007	0,0182	0,0024
RT1-1	2	0,7812	0,0032	0,1136	0,0004	0,0057	0,0102	0,0341	0,0316	0,0006	0,0181	0,0013
RT1-1	3	0,7877	0,0039	0,1029	0,0006	0,0086	0,0107	0,0281	0,0292	0,0007	0,0254	0,0022
RT1-1	4	0,7443	0,0054	0,1188	0,0007	0,0094	0,0222	0,0295	0,0320	0,0029	0,0296	0,0052
RT1-1	5	0,6309	0,0073	0,1599	0,0012	0,0224	0,0217	0,0310	0,0458	0,0028	0,0549	0,0221
RT1-1	6	0,6027	0,0085	0,1805	0,0011	0,0223	0,0210	0,0347	0,0505	0,0024	0,0513	0,0250
RT1-1	7	0,5915	0,0101	0,1910	0,0009	0,0187	0,0232	0,0377	0,0471	0,0026	0,0449	0,0323
RT1-1	8	0,5462	0,0085	0,1919	0,0016	0,0369	0,0132	0,0252	0,0605	0,0013	0,0758	0,0389
RT1-1	9	0,4949	0,0103	0,1863	0,0022	0,0402	0,0162	0,0207	0,0540	0,0022	0,0936	0,0794
RT1-10	-1	0,7166	0,0014	0,1534	0,0002	0,0037	0,0091	0,0482	0,0476	0,0005	0,0111	0,0082
RT1-10	0	0,7122	0,0017	0,1592	0,0003	0,0039	0,0102	0,0524	0,0449	0,0005	0,0118	0,0029
RT1-10	1	0,7351	0,0019	0,1445	0,0003	0,0036	0,0097	0,0478	0,0397	0,0005	0,0113	0,0056
RT1-10	2	0,8051	0,0022	0,1045	0,0003	0,0037	0,0082	0,0339	0,0279	0,0005	0,0114	0,0023
RT1-10	3	0,8369	0,0030	0,0847	0,0004	0,0043	0,0092	0,0266	0,0214	0,0007	0,0124	0,0004
RT1-16	-1	0,7247	0,0014	0,1484	0,0002	0,0033	0,0071	0,0444	0,0514	0,0005	0,0105	0,0081
RT1-16	0	0,7282	0,0015	0,1450	0,0002	0,0027	0,0080	0,0456	0,0468	0,0004	0,0098	0,0118
RT1-16	1	0,7253	0,0025	0,1466	0,0003	0,0028	0,0097	0,0482	0,0436	0,0005	0,0110	0,0095
RT1-16	2	0,7712	0,0026	0,1173	0,0003	0,0022	0,0078	0,0388	0,0354	0,0004	0,0097	0,0143
RT1-16	3	0,8362	0,0015	0,0881	0,0002	0,0022	0,0043	0,0285	0,0281	0,0003	0,0078	0,0028
RT1-16	4	0,8208	0,0022	0,0007	0,0002	0,0016	0,0087	0,0976	0,0293	0,0010	0,0076	0,0303
RT1-16	5	0,7644	0,0028	0,1248	0,0003	0,0036	0,0117	0,0394	0,0353	0,0014	0,0131	0,0032
RT1-16	6	0,7165	0,0030	0,1461	0,0003	0,0041	0,0113	0,0472	0,0393	0,0012	0,0134	0,0176
RT1-16	7	0,6525	0,0042	0,1982	0,0003	0,0041	0,0135	0,0487	0,0394	0,0015	0,0147	0,0229
RT1-16	8	0,6696	0,0042	0,1644	0,0008	0,0121	0,0117	0,0387	0,0480	0,0016	0,0336	0,0153
RT1-16	9	0,6228	0,0053	0,1733	0,0011	0,0187	0,0124	0,0378	0,0530	0,0014	0,0463	0,0279
RT1-17	-1	0,7226	0,0013	0,1451	0,0002	0,0029	0,0072	0,0437	0,0511	0,0004	0,0098	0,0157
RT1-17	0	0,7326	0,0015	0,1478	0,0002	0,0027	0,0081	0,0469	0,0473	0,0004	0,0098	0,0027
RT1-17	1	0,7279	0,0021	0,1459	0,0003	0,0026	0,0090	0,0482	0,0441	0,0005	0,0101	0,0093
RT1-17	2	0,7757	0,0024	0,1184	0,0003	0,0019	0,0073	0,0397	0,0362	0,0004	0,0084	0,0093
RT1-17	3	0,8363	0,0017	0,0893	0,0002	0,0016	0,0048	0,0295	0,0279	0,0003	0,0067	0,0017
RT1-19	-1	0,7176	0,0016	0,1449	0,0003	0,0034	0,0104	0,0482	0,0390	0,0005	0,0121	0,0220
RT1-19	0	0,7397	0,0018	0,1449	0,0002	0,0030	0,0103	0,0495	0,0379	0,0005	0,0111	0,0011
RT1-19	1	0,7464	0,0020	0,1413	0,0002	0,0026	0,0098	0,0491	0,0369	0,0005	0,0102	0,0010
RT1-19	2	0,7865	0,0030	0,1169	0,0003	0,0024	0,0099	0,0413	0,0288	0,0005	0,0103	0,0001
RT1-19	3	0,8042	0,0031	0,1015	0,0004	0,0044	0,0094	0,0336	0,0261	0,0005	0,0144	0,0024
RT1-19	4	0,7549	0,0042	0,1216	0,0005	0,0049	0,0211	0,0364	0,0299	0,0024	0,0189	0,0052
RT1-19	5	0,6793	0,0056	0,1531	0,0007	0,0102	0,0244	0,0417	0,0378	0,0029	0,0298	0,0145
RT1-19	6	0,6458	0,0071	0,1721	0,0006	0,0099	0,0248	0,0480	0,0399	0,0024	0,0281	0,0213
RT1-19	7	0,6114	0,0101	0,2009	0,0006	0,0086	0,0296	0,0489	0,0394	0,0031	0,0272	0,0202
RT1-19	8	0,6342	0,0064	0,1780	0,0008	0,0158	0,0172	0,0433	0,0478	0,0018	0,0372	0,0175
RT1-19	9	0,5402	0,0088	0,1883	0,0019	0,0343	0,0154	0,0302	0,0572	0,0019	0,0793	0,0425
RT1-20	-1	0,7225	0,0010	0,1375	0,0002	0,0023	0,0072	0,0458	0,0415	0,0004	0,0080	0,0336
RT1-20	0	0,7434	0,0011	0,1427	0,0002	0,0023	0,0083	0,0491	0,0390	0,0004	0,0084	0,0051
RT1-20	1	0,7431	0,0013	0,1439	0,0002	0,0022	0,0083	0,0508	0,0384	0,0004	0,0082	0,0032
RT1-20	2	0,7797	0,0022	0,1173	0,0002	0,0024	0,0085	0,0419	0,0289	0,0004	0,0090	0,0095
RT1-20	3	0,7955	0,0025	0,1016	0,0004	0,0058	0,0075	0,0328	0,0277	0,0005	0,0163	0,0094

**Tabela C.7:** Continuação da Base de Dados 3

Amostra	Tipo do Grão	$SiO_2$ %	$TiO_2$ %	$Al_2O_3$ %	$MnO$ %	$MgO$ %	$CaO$ %	$Na_2O$ %	$K_2O$	$P_2O_5$ %	$Fe_2O_3$ %	Outros
RT1-3	-1	0,7347	0,0026	0,1397	0,0004	0,0058	0,0120	0,0421	0,0399	0,0008	0,0184	0,0036
RT1-3	0	0,7302	0,0030	0,1399	0,0004	0,0057	0,0117	0,0430	0,0393	0,0008	0,0186	0,0074
RT1-3	1	0,7303	0,0029	0,1352	0,0004	0,0052	0,0113	0,0425	0,0378	0,0007	0,0173	0,0164
RT1-3	2	0,7790	0,0033	0,1160	0,0004	0,0051	0,0106	0,0356	0,0316	0,0006	0,0169	0,0009
RT1-3	3	0,7897	0,0041	0,1020	0,0006	0,0069	0,0125	0,0288	0,0275	0,0008	0,0221	0,0050
RT1-7A	-1	0,7111	0,0022	0,1586	0,0003	0,0051	0,0143	0,0547	0,0353	0,0007	0,0150	0,0027
RT1-7A	0	0,7063	0,0026	0,1614	0,0003	0,0053	0,0152	0,0572	0,0344	0,0008	0,0155	0,0010
RT1-7A	1	0,7150	0,0031	0,1570	0,0003	0,0046	0,0150	0,0567	0,0328	0,0008	0,0144	0,0003
RT1-7A	2	0,7891	0,0030	0,1159	0,0002	0,0033	0,0113	0,0419	0,0239	0,0005	0,0106	0,0000
RT1-7A	3	0,8243	0,0030	0,0924	0,0003	0,0036	0,0101	0,0327	0,0193	0,0007	0,0107	0,0029
RT1-7A	4	0,7718	0,0046	0,1137	0,0004	0,0047	0,0213	0,0369	0,0229	0,0028	0,0170	0,0039
RT1-7A	5	0,6806	0,0058	0,1530	0,0006	0,0119	0,0269	0,0420	0,0330	0,0032	0,0302	0,0128
RT1-7A	6	0,6421	0,0076	0,1748	0,0006	0,0124	0,0278	0,0467	0,0378	0,0030	0,0308	0,0164
RT1-7A	7	0,6370	0,0085	0,1852	0,0005	0,0106	0,0265	0,0507	0,0379	0,0029	0,0271	0,0131
RT1-7A	8	0,5854	0,0074	0,1893	0,0010	0,0306	0,0178	0,0359	0,0511	0,0018	0,0533	0,0264
RT1-7A	9	0,5452	0,0088	0,1906	0,0015	0,0412	0,0174	0,0319	0,0551	0,0017	0,0732	0,0334
RT1-7B	-1	0,7111	0,0025	0,1562	0,0003	0,0063	0,0132	0,0531	0,0359	0,0007	0,0159	0,0048
RT1-7B	0	0,6445	0,0026	0,1421	0,0003	0,0064	0,0163	0,0633	0,0315	0,0008	0,0144	0,0778
RT1-7B	1	0,7090	0,0032	0,1566	0,0003	0,0052	0,0152	0,0560	0,0325	0,0008	0,0149	0,0063
RT1-7B	2	0,7847	0,0033	0,1159	0,0003	0,0041	0,0118	0,0420	0,0239	0,0006	0,0116	0,0018
RT1-7B	3	0,7856	0,0032	0,0889	0,0003	0,0043	0,0103	0,0330	0,0188	0,0008	0,0117	0,0431
RT1-8	-1	0,7283	0,0021	0,1424	0,0003	0,0054	0,0110	0,0532	0,0271	0,0006	0,0132	0,0164
RT1-8	0	0,7076	0,0018	0,1565	0,0003	0,0046	0,0112	0,0576	0,0333	0,0006	0,0122	0,0143
RT1-8	1	0,7171	0,0019	0,1529	0,0002	0,0045	0,0107	0,0573	0,0311	0,0006	0,0115	0,0122
RT1-8	2	0,7844	0,0021	0,1151	0,0002	0,0042	0,0086	0,0433	0,0230	0,0005	0,0100	0,0086
RT1-8	3	0,8137	0,0027	0,0970	0,0003	0,0063	0,0088	0,0346	0,0196	0,0008	0,0134	0,0028
RT1-8	4	0,7602	0,0046	0,1206	0,0004	0,0076	0,0195	0,0393	0,0239	0,0028	0,0204	0,0007
RT1-8	5	0,6630	0,0062	0,1582	0,0007	0,0163	0,0237	0,0427	0,0344	0,0031	0,0351	0,0166
RT1-8	6	0,6414	0,0070	0,1775	0,0006	0,0145	0,0224	0,0501	0,0377	0,0026	0,0310	0,0152
RT1-8	7	0,6097	0,0085	0,2029	0,0006	0,0128	0,0245	0,0518	0,0361	0,0032	0,0296	0,0203
RT1-8	8	0,5758	0,0073	0,1905	0,0011	0,0341	0,0170	0,0347	0,0516	0,0018	0,0563	0,0298
RT1-8	9	0,5185	0,0089	0,1925	0,0016	0,0469	0,0159	0,0282	0,0568	0,0019	0,0780	0,0508
RT1-9	2	0,7501	0,0030	0,1114	0,0007	0,0200	0,0059	0,0256	0,0352	0,0004	0,0338	0,0139
RT1-9	3	0,7850	0,0029	0,0997	0,0006	0,0150	0,0066	0,0254	0,0293	0,0005	0,0269	0,0081
RT1-9	4	0,7525	0,0047	0,1169	0,0007	0,0120	0,0174	0,0303	0,0307	0,0025	0,0291	0,0032
RT1-9	5	0,6375	0,0067	0,1608	0,0011	0,0261	0,0207	0,0330	0,0440	0,0026	0,0522	0,0153
RT1-9	6	0,6237	0,0092	0,1747	0,0009	0,0183	0,0289	0,0421	0,0407	0,0035	0,0410	0,0170
RT1-9	8	0,5569	0,0067	0,1859	0,0014	0,0434	0,0117	0,0267	0,0584	0,0013	0,0704	0,0372
RT1-9	9	0,5013	0,0079	0,1965	0,0018	0,0566	0,0110	0,0198	0,0620	0,0014	0,0904	0,0513

# Referências Bibliográficas

- Aitchison(1982)** J Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society, Series B*, 44, 139–177. Citado na pág. 2
- Aitchison(1983)** J Aitchison. Principal component analysis of compositional data. *Biometrika*, 70, 57–65. Citado na pág. 2
- Aitchison(1984)** J Aitchison. Reducing the dimensionality of compositional data sets. *Journal of the International Association for Mathematical Geology*, 16, 617–635. Citado na pág. 2
- Aitchison(1985)** J Aitchison. A general class of distributions on the simplex. *Journal of the Royal Statistical Society, Series B*, 47, 136–146. Citado na pág. 2, 6
- Aitchison(1986)** J Aitchison. *The Statistical Analysis of Compositional Data*. Chapman and Hall London. Citado na pág. 2
- Aitchison e Shen(1980)** J Aitchison e S Shen. Logistic-normal distributions: Some properties and uses. *Biometrika*, 67, 261–272. Citado na pág. 2
- Aitchison et al.(2002)** J Aitchison, C Barceló-Vidal, JJ Egozcue e V Pawlowsky-Glahn. A concise guide for the algebraic-geometric structure of the simplex, the sample space for compositional data analysis. *Proceedings of IAMG*, 2, 387–392. Citado na pág. 4, 6
- Baxter et al.(2005)** MJ Baxter, CC Beardah, HEM Cool e CM Jackson. Compositional data analysis of some alkaline glasses. *Mathematical Geology*, 37, 183–196. Citado na pág. 3
- Butler(1979)** J Butler. Effects of closure on the measures of similarity between samples. *Mathematical Geology*, 11, 431–440. Citado na pág. 2
- Chayes(1956)** F Chayes. *Petrographic Modal Analysis: an Elementary Statistical Appraisal*. Wiley New York. Citado na pág. 2
- Chayes(1960)** F Chayes. On correlation between variables of constant sum. *Journal of Geophysical Research*, 65, 4185–4193. Citado na pág. 2
- Chayes(1971)** F Chayes. *Ratio Correlation: a Manual for Students of Petrology and Geochemistry*. University of Chicago Press. Citado na pág. 2
- Chayes e Kruskal(1966)** F Chayes e W Kruskal. An approximate statistical test for correlations between proportions. *The Journal of Geology*, 74, 692–702. Citado na pág. 2
- Chayes e Trochimczyk(1978)** F Chayes e J Trochimczyk. An effect of closure on the structure of principal components. *Journal of the International Association for Mathematical Geology*, 10, 323–333. Citado na pág. 2

- Darroch e James(1974)** JN Darroch e IR James. F-independence and null correlation of continuous, bounded-sum, positive variables. *Journal of the Royal Statistical Society, Series B*, 36, 467–483. Citado na pág. 2
- Feng et al.(2014)** C Feng, H Wang, N Lu, T Chen, H He, Y Lu e X Tu. Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, 26, 105–109. Citado na pág. 5
- Gentleman et al.(1997)** Robert Gentleman, Ross Ihaka, D Bates et al. The r project for statistical computing. *R home web site: http://www.r-project.org*. Citado na pág. 9
- Gentleman et al.(2013)** Robert Gentleman, Ross Ihaka, D Bates et al. Analyzing compositional data with r (supplementary material). Site: <http://www.stat.boogaart.de/compositionsRBook/>. Citado na pág. 116
- Geweke(1992)** J Geweke. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. *Computing Science and Statistics: Proceedings of the 23rd Symposium*, 571–578. Citado na pág. 23
- Grantham e Velbel(1988)** J Grantham e M A Velbel. The influence of climate and topography on rock-fragment abundance in modern fluvial sands of the southern blue ridge mountains, north carolina. *Journal of Sedimentary Research*, 58, 219–227. Citado na pág. 13, 115
- Kuhn et al.(2000)** T Kuhn, H Burger, D Castradori e P Halbach. Volcanic and hydrothermal history of ridge segments near the rodrigues triple junction (central indian ocean) deduced from sediment geochemistry. *Marine Geology*, 169, 391–409. Citado na pág. 10, 112
- Luc(1986)** Devroye Luc. *Non-Uniform Random Variate Generation*. Springer. Citado na pág. 22
- McDonald(2009)** J McDonald. *Handbook of Biological Statistics*, volume 2. Sparky House Publishing Baltimore, MD. Citado na pág. 5
- Mosimann(1962)** J Mosimann. On the compound multinomial distribution, the multivariate  $\beta$ -distribution, and correlations among proportions. *Biometrika*, 49, 65–82. Citado na pág. 2
- Neocleous et al.(2011)** T Neocleous, C Aitken e G Zadora. Transformations for compositional data with zeros with an application to forensic evidence evaluation. *Chemometrics and Intelligent Laboratory Systems*, 109, 77–85. Citado na pág. 7
- Pearson(1896)** K Pearson. Mathematical contributions to the theory of evolution.—on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, 60, 489–498. Citado na pág. 2
- Tolosana-Delgado e von Eynatten(2010)** R Tolosana-Delgado e H von Eynatten. Simplifying compositional multiple regression: Application to grain size controls on sediment geochemistry. *Computers & Geosciences*, 36, 577–589. Citado na pág. 16, 116
- Van den Boogaart e Tolosana-Delgado(2013)** K Van den Boogaart e R Tolosana-Delgado. *Analyzing Compositional Data with R*. Springer. Citado na pág. 4, 9, 116

**Wang et al.(2007)** H Wang, Qg Liu, H Mok, Li Fu e W Tse. A hyperspherical transformation forecasting model for compositional data. *European Journal of Operational Research*, 179, 459–468. Citado na pág. 6