UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GEOCIÊNCIAS

Abordagem estatística aplicada à análise geoquímica do Alogrupo Guaritas, Bacia do Camaquã – RS, Brasil.

Ariane Kravczyk Bernardes

ORIENTADORA:
Profª. Drª. Ana Maria Pimentel Mizusaki

COMISSÃO EXAMINADORA:

Prof. Dr. Giovani Matte Cioccari (UFPEL)
Profa. Dra. Cassiana Michelin (UFRGS)
Prof. Dr. Rogério Roque Rubert (UFMT)

Dissertação de Mestrado apresentada como requisito parcial para obtenção do título de Mestre em Geociências

Porto Alegre, 2021

Kravczyk Bernardes , Ariane
Abordagem estatística aplicada à análise geoquímica do Alogrupo Guaritas, Bacia do Camaquã - RS, Brasil. / Ariane Kravczyk Bernardes . -- 2021.
104 f.
Orientadora: ANA MARIA PIMENTEL MIZUSAKI.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Geociências, Programa de Pós-Graduação em Geociências, Porto Alegre, BR-RS, 2021.

1. Estratigrafia química. 2. Estatística. I. PIMENTEL MIZUSAKI, ANA MARIA, orient. II. Título.

LISTA DE FIGURAS

Figura 1 – Contexto geológica da Bacia do Camaquã inserida no ESRG (modificado de Chemale Jn., 2000).

Figura 2 – Localização e mapa geológico do Alogrupo Guaritas (modificado de Paim *et al. 2000).*

Figura 3 – Estágios tectônicos da Bacia do Camaquã e sua divisão em formações. (modificado de Chemale Jr. (2000).

Figura 4 – Seção estratigráfica correspondente ao local de estudo no qual foram realizadas as coletas das amostras.

LISTA DE TABELAS

Tabela 1 – Principais características que diferenciam dados brutos de dados estatisticamente tratados, de acordo com Andriotti (2003).

LISTA DE SIGLAS E DE ABREVIAÇÕES

BC – Bacia do Camaquã
ESRG – Escudo Sul Rio Grandense
ACP – Análise das Componentes Principais
CP – Componentes Principais
ICP - Inductively Coupled Plasma

Agradecimentos

Agradeço ao Programa de Pós-Graduação em Geociências da Universidade Federal do Rio Grande do Sul e a toda sua equipe pelo apoio prestado, aos professores da anterior e atual coordenação, prof. Rommulo e profa. Maria de Fatima, respectivamente. Agradeço ao suporte sempre atencioso prestado pelo Roberto, ao colega e amigo Deleon pelas palavras de incentivo, igualmente a minha amiga Zélia pelo carinho e apoio. Gostaria de agradecer a colega Patrícia pelo auxílio com as imagens e a todos e todas colegas que de um algum modo, por meio de suas pesquisas, contribuíram para que este trabalho pudesse acontecer.

Faço um agradecimento mais do que especial para a profa. Ana Maria Pimentel Mizusaki, minha orientadora, pelo inestimado apoio e extrema paciência nas horas difíceis, que infelizmente não foram poucas nesses últimos tempos. Ana, muito obrigada por ter me incentivado até o final, quando eu mesma já não acreditava que seria possível.

Agradeço a todos os amigos, familiares, meus pais, Mariane e Ulisses e a todas as pessoas que de maneira geral nos incentivam a seguir estudando e buscando aprimorar o conhecimento. Agradeço, é claro, aos meus amados Hami e Luna pela companhia calma e tranquilizadora durante as horas de estudo.

Novamente, faço um agradecimento mais do que especial ao meu amado amigo e companheiro de vida, Álvaro. Agradeço por todo apoio e ajuda incondicionais que você me dá e por saber que sempre poderei contar com você. A caminhada da vida ganha um colorido especial ao seu lado.

Por fim, agradeço a minha querida e amada avó, Maria Magdalena, a quem dedico este trabalho, que com muito esforço pude finalizar. *Vó*, você foi e sempre será como uma mãe para mim, não importa onde estejas, estará sempre em meu coração. Serei eternamente grata por todas as coisas bonitas da vida que você me mostrou e por ter me ensinado tão bem a amar e a respeitar toda a natureza deste nosso incrível planeta.

SUMÁRIO

Lista de figuras....03
Lista de quadros....04
Lista de siglas e abreviaturas....05
Agradecimentos....06
Sumário....07
Resumo....08
Abstract....09
Estrutura da dissertação....10
1. Introdução....11
    1.1 Justificativa....14
    1.2 Objetivos....21
2. Área de estudo....22
    2.1 Dados Analisados....26
3. Estratigrafia química....28
4. Geoestatística....30
5. Metodologia....31
    5.1 Banco de dados geoquímicos....31
    5.2 Métodos estatísticos....33
    5.2.1 Retas de Correlação....33
    5.2.2 Significância estatística....34
    5.2.3 Análise das Componentes Multivariadas....37
6. Resultados obtidos....41
    6.1 Carta de Submissão....41
    6.2 Manuscrito....42
7. Referências....74
8. Considerações Finais....77
Apêndice....79
Anexos....99

RESUMO

Nesta dissertação foram empregados métodos estatísticos para a análise de dados geoquímicos obtidos por meio da técnica de espectrometria de massa por fusão da amostra com plasma acoplado indutivelmente (FUS-ICP: Fusion-Inductively Coupled Plasma). A utilização de técnicas estatísticas amplia as condições de análise e interpretação da informação contida em bancos de dados. No contexto das Geociências, especificamente neste estudo, foi realizado o tratamento estatístico de dados geoquímicos. Esse tipo de abordagem permitiu estabelecer correlações de elementos químicos maiores, menores e traços, assim como critérios de variabilidade da composição química (com significância estatística) para as rochas do Alogrupo Guaritas na bacia do Camaquã, estado do Rio Grande do Sul. Além de contribuir com a consolidação do conhecimento para esta região, o foco deste estudo consiste na abordagem estatística para análise dos dados geoquímicos disponíveis para o Alogrupo Guaritas, provenientes do trabalho de Denalle (2013). Foram aplicados os seguintes métodos estatísticos: reta de correlação por aproximação de pontos de uma ou mais variáveis pelo método dos mínimos quadrados, com cálculo da significância estatística da correlação estabelecida para os pontos amostrados e o método da Análise de Componentes Principais – ACP. A abordagem estatística tem como objetivo a otimização dos dados disponíveis para o Alogrupo Guaritas, com vistas ao refinamento da informação e de suas interpretações possíveis, que melhor estabeleçam as caraterísticas composicionais e relações de gênese para estas rochas. Nos pontos amostrados há variações sutis na composição química, porém sem variações macroscópicas significativas observadas. O estabelecimento de retas de correlação auxilia na compreensão das relações específicas da composição de um elemento químico, que pode ser, por exemplo, um elemento maior presente em abundância em relação a variação de um ou dois elementos menores ou traço. Enquanto que o método de Análise das Componentes Principais – ACP ou PCA (do inglês, Principal Component Alalysis) contribui para o entendimento da influência de quais conjuntos de elementos químicos são determinantes na diferenciação dos tipos litológicos. Essa técnica se mostra vantajosa para a análise de extensos bancos de dados, na medida em que reduz a quantidade excessiva de dados, eliminando sobreposições, porém sem perda significante de informação. A ACP aponta para escolhas das formas mais representativas dos dados por meio de combinações lineares das variáveis originais. Esta técnica confere vantagens no gerenciamento de dados geoquímicos, dentre outras áreas de pesquisa em geociências, pois pode ser utilizada de forma análoga para diversas bases de dados disponíveis. O emprego de métodos estatísticos neste trabalho para o Alogrupo Guaritas mostrou-se eficaz, mostrando relações composicionais entre elementos maiores, formadores das rochas, como SI, Ca, Al e Mg e elementos menores, como Mn, Zn, Ti, Rb e Sr. Portanto, os resultados preliminares apontam que o uso de técnicas de análise estatística para bases de dados geoquímicos consiste em uma ferramenta efetiva, usada de forma adicional para interpretação de dados geoquímicos, como suporte às técnicas analíticas tradicionalmente empregadas em estratigrafia química.

Palavras-chave: Análise Estatística, Componentes Principais, Estratigrafia Química, Ordoviciano, Aloestratigrafia.

ABSTRACT

In this dissertation, statistical methods were used for the analysis of geochemical data obtained through the technique of sample fusion mass spectrometry with inductively coupled plasma (FUS-ICP: Fusion-Inductively Coupled Plasma). The use of statistics expands as conditions for analysis and interpretation of information contained in databases. In the context of Geosciences, specifically in this study, the statistical treatment of geochemical data was carried out. This type of common approach establishes correlations of major, minor and trace chemical elements, as well as criteria of chemical composition variability (with statistical significance) for the rocks of the Guaritas Allogroup in the Camaquã basin, state of Rio Grande do Sul. With the consolidation of knowledge for this region, the focus of this study is the statistical approach for analyzing the geochemical data available for the Alogrupo Guaritas, from the work of Denalle (2013). The following statistical methods were used: resumption of the relationship by approximating the points of one or more variables through the method of collected data, with calculation of the statistical significance of the estimate for the sampled points and the Principal Component Analysis - PCA method. The statistical approach aims to optimize the data available for the Allogroup Guaritas, with a view to refine the information and its possible interpretations, which better establish the compositional characteristics and genesis relationships for these rocks. At the sampled points, subtle variations in chemical composition, but no macroscopic variations observed. The establishment of correlation lines helps to understand the specific relationships of a chemical element, which can be, for example, a larger element present in relation to the variation of one or two smaller elements or trace. While the Principal Component Analysis method contributes to the understanding of the influence of which sets of chemical elements are decisive in the differentiation of lithological types. This technique proves to be advantageous for an analysis of extensive databases, as it reduces the excessive amount of data, eliminating overlaps, but without significant loss of information. The PCA researches to choose the most representative forms of the data through linear patterns of the original variables. This technique offers advantages in the management of geochemical data, among other areas of research in geosciences, as it can be used in an analogous way for several available databases. The use of statistical methods in this work for the Guaritas Allogroup proved to be effective, showing the compositional relationships between major elements, forming rocks, such as SI, Ca, Al and Mg, and minor elements, such as Mn, Zn, Ti, Rb and Sr Therefore, preliminary results indicate that the use of statistical analysis techniques for geochemical databases is an effective tool, used additionally for the interpretation of geochemical data, in support of the analytical techniques traditionally employed in chemical stratigraphy.

Keywords: Statistical Analysis, Principal Components, Chemical Stratigraphy, Ordovician, Allostratigraphy.

ESTRUTURA DA DISSERTAÇÃO

Esta dissertação de mestrado está organizada sob a forma de um capítulo introdutório ao manuscrito "**Métodos estatísticos aplicados à análise geoquímica de dados do Alogrupo Guaritas, Bacia do Camaquã – RS, Brasil**", submetido a revista Journal of South American Earth Sciences (Qualis CAPES A3).

Dessa forma, a estruturação do trabalho compreende os seguintes títulos:

- Introdução: contendo uma apresentação do tema e os objetivos da pesquisa;
- Localização da área de estudo: Contendo um resumo da evolução litoestratigráfica, tectônica, estratigrafia química;
- Estratigrafia Química;
- Geoestatística;
- Metodologia de estudo: descrevendo os métodos e materiais utilizados para o estudo; Banco de dados e métodos estatísticos;
- Resultados obtidos: síntese do manuscrito; carta de submissão para revista técnica e o artigo apresentado no seu formato integral;
- Referências bibliográficas;
- Considerações Finais.

1. Introdução

A Bacia do Camaquã (BC) localizada no Escudo Sul-rio-grandense (ESRG), caracteriza-se pela complexidade geológica, posto que seu registro estratigráfico apresenta ampla variedade de tipos litológicos. Estão presentes desde rochas sedimentares de origem marinha, aluvial e eólica até rochas vulcânicas e vulcanoclásticas de composições variadas. Eventos tectônicos intensos de forma sucessiva foram responsáveis por sua deformação, como também por processos de erosão e deposição das rochas sedimentares da BC. Esse cenário contribuiu para um ambiente geológico de difícil caracterização como vemos em Paim *et al* (2000). A (Fig. 1) mostra o ambiente geológico do ESRG no qual está inserida a BC. A porção central do ESRG é ocupada pelo domínio Santana da Boa Vista, limitado a oeste pela sutura de Caçapava e a leste pela zona de cisalhamento Dorsal de Canguçu.

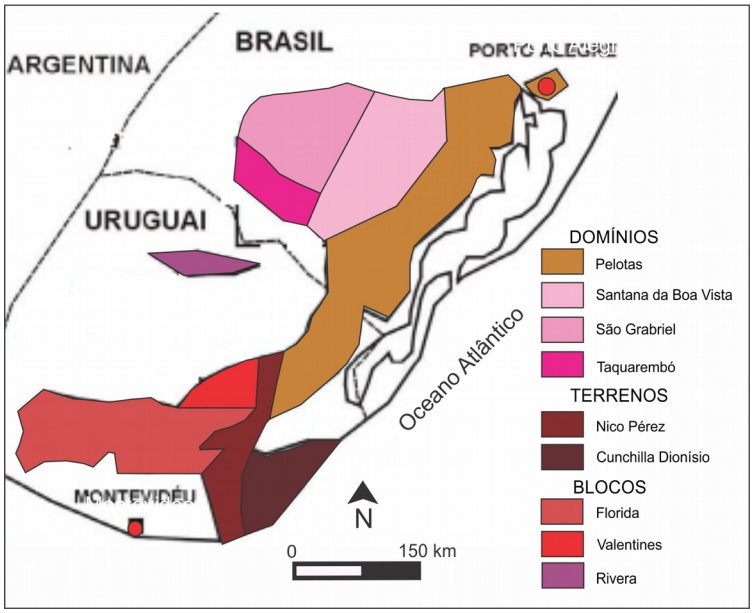


Figura 1 - Contexto geológica da Bacia do Camaquã inserida no ESRG (modificado de Chemale Jn., 2000).

Os primeiros estudos da Bacia do Camaquã datam do começo da década de 30 (Carvalho, 1932). Desde então, numerosos trabalhos foram realizados ao longo das décadas seguintes fornecendo certa abundância de interpretações e modelos. A partir das últimas décadas do século XX, por meio de uma abordagem predominantemente influenciada pela Teoria da Tectônica de Placas, as pesquisas culminaram em uma síntese acerca da evolução da Bacia do Camaquã sob a ótica de ciclos tectono-sedimentares superpostos em um mesmo *locus* deposicional (Paim *et al.,* 2014).

A BC notadamente deformada pela ação tectônica foi alvo de importantes trabalhos de campo, nos quais é fundamental a identificação das unidades rochosas *in situ,* bem como suas relações de campo. O conhecimento a respeito das formações que compõem esta bacia foi acumulado ao longo do tempo com a contribuição de diversos estudos, entre os quais, não podendo elencar a todos, daremos destaque aos desenvolvidos a partir da década de 60. As pesquisas anteriores a este período fundamentavam-se na Teoria Geossinclinal. Salienta-se, entretanto, que estes estudos contribuíram substancialmente por sua precisão e riqueza de informações dos diferentes tipos litológicos e estruturas observadas em campo.

As pesquisas realizadas entre meados dos anos 70 e 80, por sua vez, foram pautadas pela influência da tectônica de placas, dando ênfase em aspectos da sedimentologia, por meio de modelos deposicionais, pelos quais buscava-se compreender os efeitos tectônicos sofridos pela bacia. Os trabalhos desenvolvidos a partir dos anos 90 abriram caminho para uma abordagem mais ampla, envolvendo a utilização de importantes ferramentas da Estratigrafia, como a Estratigrafia de Sequências e a Estratigrafia Química. Foi por este período que a concepção de uma única “Bacia do Camaquã” deu lugar ao

entendimento de que na verdade uma série de diferentes bacias sedimentares se sucederam no tempo ocupando um mesmo *locus* deposicional.

Em Paim *et al.* (2000) temos a síntese e integração do conhecimento adquirido nas décadas anteriores. Contudo, os grupos e formações pertencentes à BC oferecem ainda desafios na busca da compreensão de sua gênese e contexto paleoambiental, sobretudo, no que se refere à geoquímica e geocronologia das formações de rochas sedimentares. Trabalhos mais recentes, como Denalle (2013), realizaram estudos de campo na BC, com vistas à aplicação da Estratigrafia Química, na busca de melhor compreensão das relações genéticas de tipos litológicos contrastantes presentes em suas formações.

Para Paim *et al.* (2000) há diversos pontos de fragilidade a respeito da BC, tais como: carência de informações quanto às relações de contato com as rochas magmáticas e o embasamento; fragilidade na definição dos episódios de deformação que afetaram as bacias. Ainda conforme Paim *et al.* (2000), são pouco abundantes os dados geocronológicos e paleontológicos, que possam definir as idades deposicionais para os estratos sedimentares, como é verificado nas suas unidades basais (Maricá e Bom Jardim) de difícil distinção.

Assim, a despeito de todo empenho nos estudos de campo realizados, o diversificado panorama do contexto geológico do que se convencionou chamar “Bacia do Camaquã” carece de mais pesquisa, não somente nos campos da sedimentologia e estratigrafia, mas também de estudos com emprego de técnicas analíticas diversificadas, entre as quais será dado, nesta dissertação, enfoque na estratigrafia química, interpretada por meio da utilização de métodos estatísticos aplicados aos dados geoquímicos.

1.1 Justificativa

A linha sucessiva de pesquisas da Bacia do Camaquã, como brevemente introduzida, evidencia a dificuldade no fechamento de dados consolidados para esta bacia. A despeito de haver importantes estudos de estratigrafia de sequências e aloestratigrafia como em Paim *et al*. (1995, 2000), assim como sobre os efeitos do tectonismo observados na BC em Chemale Jr. (2000), ainda se faz presente fortemente a necessidade de aplicação de metodologias que contribuam com dados adicionais de detalhe. Por exemplo, como as variações na composição química tanto para estratos sabidamente distintos, como também, para estratos semelhantes macroscopicamente, porém, com presença de descontinuidades entre seus contatos.

Ferramentas como geologia isotópica e geocronologia contribuem fundamentalmente para integração dos dados levantados nos estudos estratigráficos, tal qual verificamos em Borba (2006). No trabalho de Denalle (2013) verifica-se a contribuição do uso da estratigrafia química junto aos estudos de estratigrafia de sequências para o Alogrupo Guaritas. Esses trabalhos se utilizam de técnicas analíticas aprimoradas sem as quais não se poderia seguir no aprofundamento da pesquisa geológica seja para a BC ou para qualquer outro foco de estudo estratigráfico.

Nesse sentido, este trabalho apresenta uma possibilidade de suporte às técnicas já existentes, por meio de uma abordagem estatística dos dados produzidos por estas técnicas analíticas já consagradas nas pesquisas em

geociências. Dentre elas, esta dissertação destacará, em particular, a estratigrafia química. Para a qual o suporte estatístico mostra-se uma possibilidade viável e vantajosa na medida em que apresenta-se como um recurso de relativo baixo custo, capaz de fornecer interpretações de significado geológico com base em dados geoquímicos estatisticamente tratados.

A utilização da estatística vem sendo cada vez mais popularizada nas pesquisas em geociências desde meados dos anos 60, muito embora não se tenham esgotados todos os aspectos e desafios envolvidos em sua aplicação, sobretudo, em relação a utilização de dados geoquímicos. Diferentemente do que ocorre com áreas de pesquisa em geoestatística, como avaliação por amostragem de minérios de interesse em amplas zonas de mineração ou como geotecnia e cartografia, o uso de métodos estatísticos aplicados a dados geoquímicos, obtidos por meio de pesquisas em estratigrafia química, não é ainda densamente explorado.

Dessa forma, este trabalho pretende contribuir com o uso de técnicas da estatística que possibilitem o aprimoramento da informação gerada com base nos dados geoquímicos produzidos no trabalho de Denalle (2013), referentes ao Alogrupo Guaritas (Bacia do Camaquã). Cabe salientar, entretanto, que a necessidade do uso de métodos de estatística se faz presente em diversos segmentos de pesquisa em geociências e que o conjunto de técnicas estatísticas abordadas neste trabalho poderá servir de base para conjuntos de dados geoquímicos de perfil semelhante aos aqui utilizados.

Nesse sentido, o escopo principal desse trabalho será para além de fornecer interpretações em relação às variações na composição química de litologias muito semelhantes macroscopicamente do Alogrupo Guaritas, também contribuir

com o desenvolvimento de uma abordagem estatística para a utilização de dados brutos de geoquímica de elementos maiores e menores de rocha total, obtidos por meio da técnica de espectrometria de massa por fusão da amostra com plasma acoplado indutivo (FUS-ICP: Fusion-Inductively Coupled Plasma). Esse fluxo de técnicas estatísticas e seus desdobramentos por si contribuem com o conhecimento acerca das possibilidades de tratamentos de dados brutos e com o mais adequado aproveitamento dos bancos de dados em geociências de modo geral.

O conjunto de técnicas analíticas usadas em estudos de estratigrafia química podem produzir extensos bancos de dados geoquímicos a depender da quantidade de análises realizadas por amostra. Logo se pode dimensionar a complexidade existente para uma grande quantidade de dados, que muitas vezes são bastante semelhantes entre si. Assim, a simples visualização da informação fornecida pelos dados geoquímicos, que em geral se apresentam listados em extensas planilhas divididas em linhas de amostras por colunas de análises, dificilmente fornecerá interpretações rápidas e precisas da informação.

Portanto, é fundamental o manejo cada vez mais sofisticado dos métodos estatísticos com vistas a otimização da informação contida em bancos de dados geoquímicos. Uma vez que pelo não tratamento e organização adequados dos dados brutos obtidos, após análises de laboratório, vemos como consequência a subutilização da informação contida em bancos de dados. Esse fato pode acarretar dispêndio pelo alto custo das análises químicas laboratoriais, na medida em que a informação gerada por estas análises pode não ser integralmente utilizada em sua potencialidade máxima, visto que parte de um extenso banco de dados pode permanecer "ocioso" quando não submetido a métodos que incluam todos os dados nele contidos.

O aproveitamento pleno de um banco de dados pode se dar por meio de técnicas matemáticas geradoras de gráficos em que se pode observar o tipo de comportamento de uma ou mais variáveis em relação às demais. Ou seja, por meio de técnicas estatísticas nas quais possamos observar tendências de comportamentos dos dados amostrados.

Bancos de dados com grande quantidade de parâmetros e de pontos amostrados apresentam, portanto, desafios para a otimização de toda informação neles contida. Assim estudos como este buscam contribuir com o melhoramento dos fluxos de processos necessários para o tratamento estatístico de dados brutos. Entenda-se “dados brutos” como o conjunto de dados que são entregues pelo laboratório conforme o tipo de análise efetuada. São os dados fornecidos tal qual são obtidos pelo equipamento de determinada técnica analítica.

As análises por ICP, por exemplo, resultam em valores correspondentes às quantidades de elementos químicos presentes em rocha total e são apresentados no formato de matriz de dados nas quais cada coluna identifica a amostra analisada e cada linha o parâmetro mensurado, neste caso a espectrometria de massa de cada elemento químico analisado por amostra. Enquanto que os mesmos dados, após serem tratados estatisticamente, apresentar-se-ão de modo a destacar semelhanças e diferenças estatísticas, passíveis de interpretações de significado geológico.

Dados podem ser definidos como observações documentadas ou resultantes de procedimentos de medição, como os de temperatura, volume, massa, entre outros. A disponibilidade dos dados possibilita a obtenção de informações.

Enquanto que os dados brutos, por outro lado, dizem respeito aqueles que não foram submetidos a nenhum tipo de processamento. Ou seja, não passaram por métodos de análise ou algum critério previamente especificado de organização.

Na tabela 1, são apresentadas as principais diferenças entre dados geoquímicos brutos e estatisticamente tratados. Destaca-se que embora os termos “dado” e “informação” sejam usados como sinônimos, na realidade, não o são. Toda a diferenciação das características observadas nos dados após seu tratamento estatístico é que passa a constituir a informação ou conjunto de informações decorrentes do adequado tratamento dos dados. Em outras palavras, dado não é informação. Bancos de dados correspondem a conjuntos de valores numéricos ou não, obtidos por determinada técnica analítica ou determinado método de amostragem. Assim sendo, estudos que desenvolvam esse fluxo de processos necessários para conversão de dados em informação contribuem com a construção do conhecimento para uma abordagem estatística e maximização do aproveitamento dos bancos de dados existentes.

Tabela 1 - Principais características que diferenciam dados brutos de dados estatisticamente tratados, de acordo com Andriotti (2003).

DADOS BRUTOS X DADOS TRATADOS ESTATISTICAMENTE COMPARAÇÃO ENTRE DADOS GEOQUÍMICOS BRUTOS E ESTATISTICAMENTE TRATADOS	
Expressiva quantidade de informação numérica, categorizada de acordo com os pontos amostrados.	Categorização do conjunto de dados correlatos entre si, grande quantidade de informação que pode ser expressa graficamente.
Apresentação em formato de planilha na qual a visualização é feita individualmente para cada ponto amostrado com o cruzamento de cada parâmetro geoquímico analisado por vez.	Os dados podem ser expressos em diferentes tipos de gráficos de acordo com a técnica estatística empregada. A informação pode ser segmentada ou agrupada de acordo com parâmetro geoquímico de interesse.
Subutilização de bancos de dados extensos com consequente redução da precisão da informação armazenada.	Otimização do uso do conteúdo contido em vastos bancos de dados, possível correlação com outros bancos de dados semelhantes de um ou mais objetos de estudo.
Baixa ou escassa diferenciação entre dados redundantes ou indefinidos.	Identificação e seleção por níveis de significância estatística dos dados e de acordo com a necessidade do estudo.

Bacias petrolíferas, por exemplo, apresentam extensos bancos de dados, situação na qual o alto custo da sondagem para análise de novas amostras evidencia a relevância do tratamento adequado dos dados, seja para análises geoquímicas ou geofísicas da bacia. Isso porque o mais eficiente tratamento dos dados poderá fornecer informação com significância estatística e de significado geológico, tornando o efeito de um banco de dados mais eficiente em termos da informação, que pode ser gerada a partir dele com possibilidade até mesmo de diminuição de custos pela redução da necessidade de análises adicionais ou de novos pontos de amostragem.

Este trabalho busca, portanto, contribuir em dois segmentos nos quais se faz necessário o aprimoramento do conhecimento. Sendo o primeiro deles a pesquisa acerca da Bacia do Camaquã (Alogrupo Guaritas) da qual, como abordado anteriormente, ainda remanescem dúvidas e imprecisões. O segundo segmento corresponde a própria metodologia utilizada neste trabalho a ser empregada para o melhoramento da informação que pode ser extraída a partir nos dados geoquímicos disponíveis oriundos do trabalho de Denalle (2013), para o Alogrupo Guaritas (Bacia do Camaquã).

Essa metodologia trata-se da abordagem estatística que será dada com a finalidade de extrair conteúdo adicional proveniente do banco de dados utilizado neste trabalho. Como mencionado, o desenvolvimento e adaptação dos métodos estatísticos em geociências também representa um segmento importante a ser explorado. Sobretudo, como suporte às metodologias já amplamente empregadas nas pesquisas em estratigrafia química. Nesse sentido, este trabalho espera somar em alguma medida conhecimentos relevantes para o crescente melhoramento da pesquisa desenvolvida nos dois segmentos apresentados acima.

1.2 Objetivos

Este trabalho realizará o refinamento do banco de dados geoquímicos de elementos maiores, menores e traço, apresentado em Denalle (2013), para o Alogrupo Guaritas (Ordoviciano) da Bacia do Camaquã no Escudo Sul-rio-grandense, RS – Brasil. Para tal, serão aplicados métodos estatísticos para o tratamento dos dados geoquímicos disponíveis, com vistas a corroboração das interpretações efetuadas pelos estudos de estratigrafia química e geração de interpretações adicionais sobre as relações composicionais das litologias do Alogrupo Guaritas.

A abordagem estatística se dará por meio de dois principais conjuntos de métodos, a saber:

- Reta de correlação por aproximação de pontos de uma ou mais variáveis pelo método dos mínimos quadrados, com cálculo da significância estatística da correlação estabelecida para os pontos amostrados;

- Análise de Componentes Principais – ACP, método em que se efetua o procedimento matemático de transformação ortogonal de um conjunto de observações de variáveis, possivelmente correlacionadas, num conjunto de valores de variáveis linearmente não correlacionadas, chamadas de *componentes principais*. Essas técnicas serão detalhadamente apresentadas nesta dissertação no capítulo de metodologia.

Assim, este estudo tem como meta mais ampla somar conhecimento às pesquisas acerca da BC, especificamente, a respeito das unidades litológicas correspondentes ao Ordoviciano da BC, que neste estudo são referenciadas como Alogrupo Guaritas.

Como meta específica este estudo desenvolverá o fluxo de etapas necessárias para o tratamento dos dados e aplicará técnicas estatísticas existentes, efetuando as devidas adaptações para seu uso no contexto dos dados geoquímicos de rochas sedimentares.

2. Área de Estudo

O Alogrupo Guaritas localiza-se na região central do estado do Rio Grande do Sul, sobre o Escudo Sul Rio-Grandense, entre os paralelos -54°00’00” e -53°00’00” de latitude e os meridianos -30°00’00” e -31°00’00” de longitude, ao sul do município de Caçapava do Sul e a oeste-noroeste do município de Santana da Boa Vista, estendendo-se para norte até a região de Encruzilhada do Sul e distando cerca de 250 km de Porto Alegre pela rodovia BR-290 (Fig. 3).

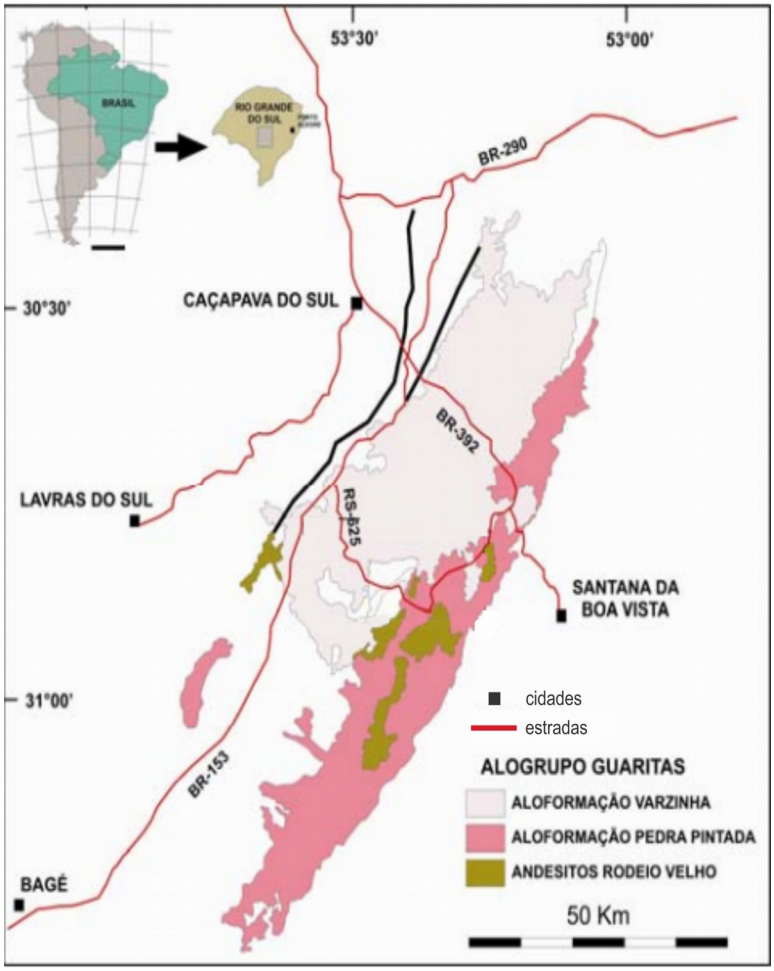


Figura 2 - Localização e mapa geológico do Alogrupo Guaritas (modificado de Paim *et al.* 2000).

O Alogrupo Guaritas é composto por rochas siliciclásticas em sua maioria arenosas e conglomeráticas, interpretadas como sendo de origem aluvial e eólica. Apresenta uma importante relação com eventos geológicos marcantes do sul do Brasil, uma vez que representa o encerramento da sedimentação na BC de tipo rifte, durante o Cambriano, em ambiente no qual foi posteriormente recoberto por ampla sedimentação que compõem a bacia intracratônica do Paraná (Ordoviciano).

Considerado como o último estágio de deposição na BC, o Alogrupo Guaritas, aflora em uma faixa alongada na região central do Escudo Sul Rio-Grandense. Ele é formado por duas grandes sequências: Pedra Pintada (base) e Varzinha (topo). Datações radiométricas (U/Pb) em zircões provenientes das rochas vulcânicas da base da sequência Pedra Pintada sugerem uma idade de 470 ± 19 Ma para o evento magmático, portanto uma provável idade ordoviciana para a sequência (Maraschin *et al*., 2007).

À BC é proposta a divisão nos alogrupos Maricá, Bom Jardim, Cerro do Bugio, Santa Bárbara e Guaritas (Borba, 2006)**,** sendo esses últimos grupos, Santa Bárbara e Guaritas representativos das últimas unidades da Bacia do Camaquã, formadas em um ambiente tectônico dominantemente transtracional, com intervalos de idade deposicional ainda controversos (Oliveira, 2012).

As litologias estão inseridas na fase final de preenchimento da Bacia do Camaquã, com idades aferidas de 470 ± 19 Ma (Ordoviciano) para as rochas vulcânicas da base do Alogrupo Guaritas, obtida pela datação de um grão de zircão pelo método U/PB, Hartmann et al. (1998). Foram também efetuadas datações pelo método K/Ar em ilitas autigênicas contidas em estratos que se situam logo acima das rochas vulcânicas da base da Aloformação Pedra Pintada, propostas por Maraschin et al. (2010), indicando idades entre 473.7 ± 9.4 e 521.7 ± 10.3 Ma (Cambriano Inferior ao Ordoviciano).

Os estudos de tectônica e de arquitetura da bacia caracterizam os eventos e distinguem as unidades litoestratigráficas para a BC neles é feito o uso da nomenclatura primeiramente descrita para as formações da Bacia do Camaquã (Fig. 4). Nesta dissertação, entretanto, se utilizará a nomenclatura

dada pela Aloestratigrafia para a denominação das unidades litológicas da BC. Entretanto, cabe destacar que ambas as terminologias empregadas tanto da estratigrafia convencional, como da aloestratigrafia não são necessariamente concorrentes e sua utilização depende do enfoque pretendido em cada estudo.

A aloestratigrafia trata-se de uma classificação que considera as unidades estratigráficas delimitadas por superfícies de descontinuidade que envolvem estas unidades, permitindo seu mapeamento geológico. Uma vez que, enquanto na litoestratigrafia, o critério principal é a identidade ou a associação litológica, na aloestratigrafia pode haver rochas de mais de um tipo de ambiente sedimentar, portanto, litologicamente diferenciadas, mas balizadas por superfícies de não-deposição ou de erosão que as correlacionam lateral e temporalmente.

Essa perspectiva dada pela aloestratigrafia se coaduna adequadamente à escolha da estratigrafia química usada para estudo e caracterização das litologias do Alogrupo Guaritas e suas aloformações e alomembros, sendo que a distinção da composição química de elementos maiores, menores e traços pode estabelecer relações genéticas entre litologias aparentemente distintas. Assim, os objetivos deste estudo coadunam-se adequadamente à utilização da aloestratigrafia, sendo que o banco de dados geoquímicos, que será analisado por meio de ferramentas da estatística, auxilia à pesquisa em estratigrafia química aplicável às unidades aloestratigráficas.

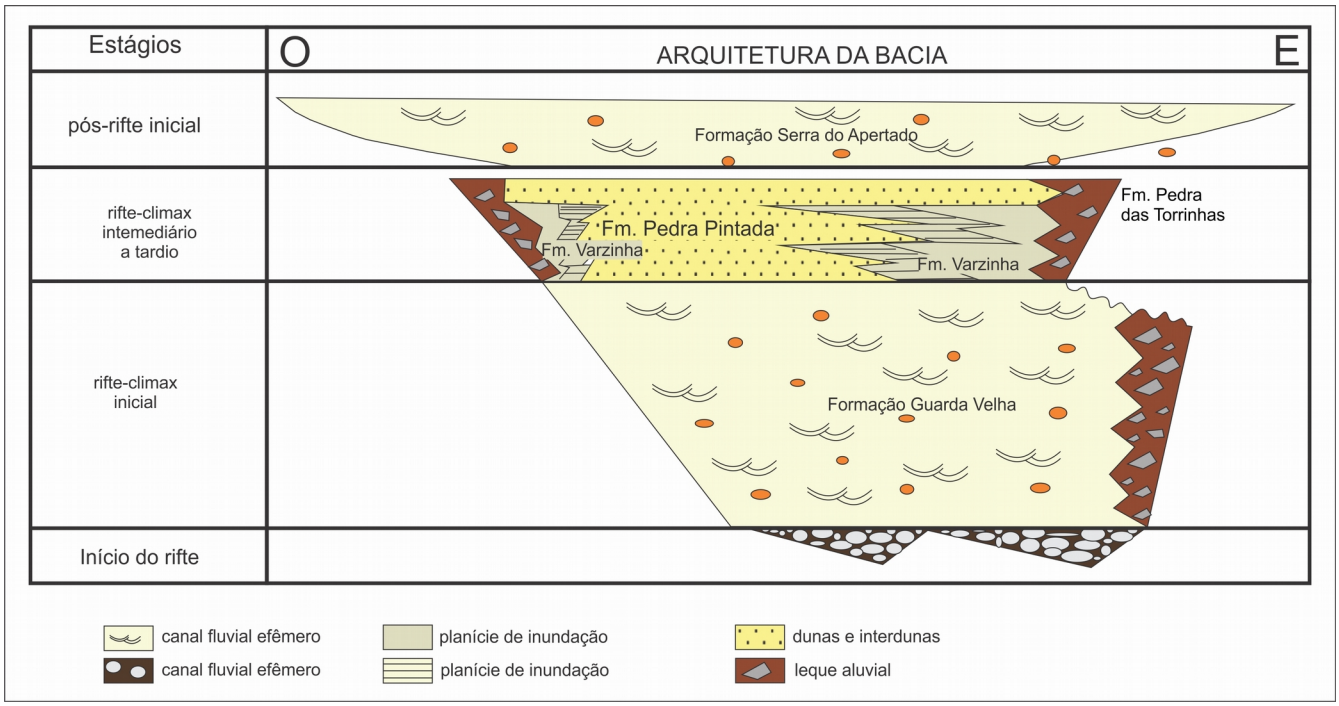


Figura 3 - Estágios tectônicos da Bacia do Camaquã e sua divisão em formações. (modificado de Chemale Jr. (2000).

2.1 Dados Analisados

Neste trabalho foi o utilizado o banco de dados provenientes do trabalho de Denalle (2013), já mencionado, que segue anexo a esta dissertação. O banco de dados é constituído de 26 pontos amostrados nos quais foram realizadas análises para 57 elementos químicos, entre elementos maiores, menores e traço. A técnica analítica utilizada para detecção dos elementos foi a espectrometria de massa por fusão da amostra com plasma acoplado indutivelmente (FUS-ICP: Fusion-Inductively Coupled Plasma). As análises químicas em rocha total para elementos maiores, menores e traço foram preparadas no Activation Laboratories Ltd. (Laboratório *Actlabs*) com sede no Canadá. As amostras de rocha foram extraídas com os adequados

procedimentos para análise química nos 26 pontos amostrados ao longo do perfil colunar, com cerca de 38 metros de espessura do Alogrupo Guaritas (Fig. 5).

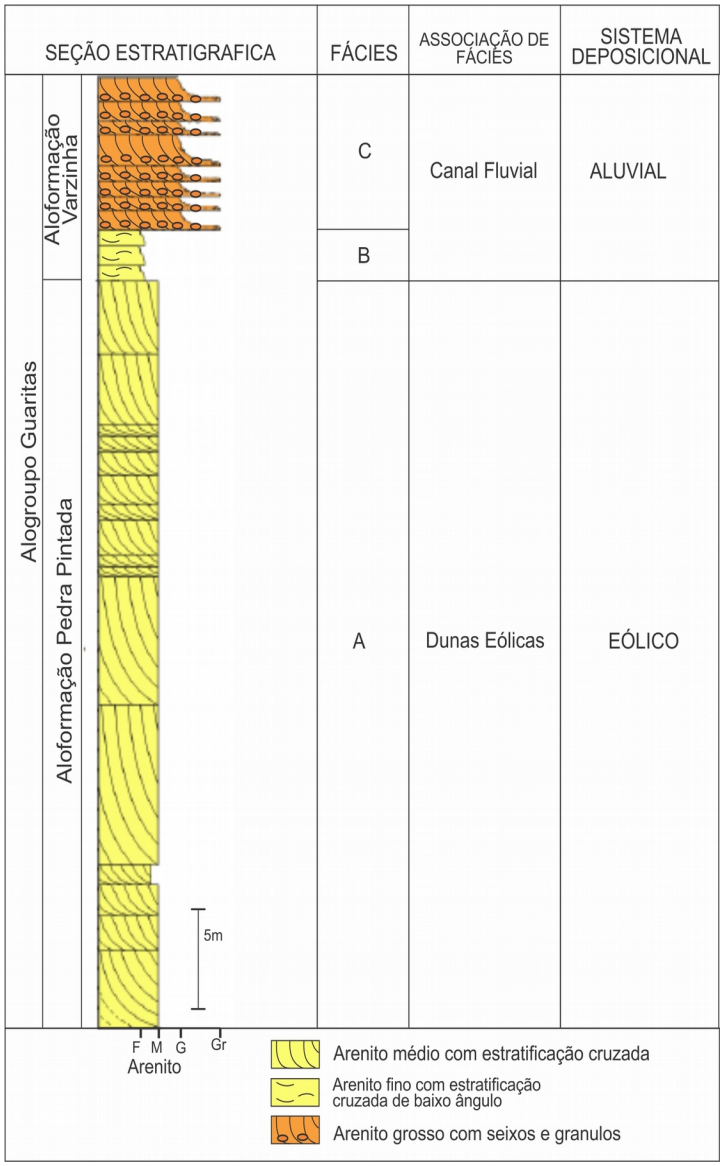


Figura 4 – Seção estratigráfica correspondente ao local de estudo no qual foram realizadas as coletas das amostras.

3. Estratigrafia Química

A geoquímica é uma ferramenta analítica aplicada que auxilia no entendimento de processos geológicos tais como reconhecimento de paleoambientes, paleoclimas, tectônica, proveniência e correlações estratigráficas em bacias sedimentares, entre outros processos. Entretanto, a caracterização de unidades deposicionais com foco na distinção e correlação com base na variação de dados químicos e o uso do termo “quimioestratigrafia” ou “estratigrafia química” para tal foram mais frequentemente empregados a partir da década de 1980, como identificamos em Ramkumar (2015).

A quimioestratigrafia abrange a identificação de características geoquímicas com base na variação espaço-temporal dos elementos químicos presentes nas rochas ou ainda na variação das razões entre esses elementos (Ramkumar, 2015 *apud* Prundeanu, 2021). Assim, a quimioestratigrafia tem sido aplicada em estudos de amplo número de bacias e unidades sedimentares pelo mundo, tais como nas Bacias Foreland Andinas (Pinto *et al*, 2018), Bacia do Karoo, África do Sul (Franchi *el al*, 2021) e Bacia do Duero na Espanha (Gómez *et al*, 2006).

A Estratigrafia Química (ou quimioestratigrafia) é uma técnica estratigráfica, que envolve a aplicação dos dados da geoquímica (orgânica ou inorgânica) com objetivo de caracterizar e correlacionar as unidades estratigráficas. Dessa forma, o estudo quimioestratigráfico pode ser aplicado em diversos ramos do conhecimento, desde as mais abordados, como geologia do petróleo para caracterização de rochas geradoras, para correlação estratigráfica,

como também na sedimentologia para compreensão da proveniência de sedimentos, estudos da variação de temperatura da massa d`água, condições biogeoquímicas marinhas de deposição de sedimentos ou mesmo para paleoclimatologia em estudos de variação isotópica de oxigênio em carbonatos ao longo do tempo geológico.

Devido ao foco de estudo desta dissertação se tratar de um banco de dados geoquímicos (inorgânicos), este trabalho relaciona-se com este segmento de pesquisa e o conhecimento de suas técnicas. Somente assim é possível gerenciar bancos de dados dessa natureza, com vistas à aplicação do tratamento estatístico que otimize as análises e interpretações possíveis.

4. Geoestatística

O termo *Geoestatística é* designado para representar a área de estudo específica em geociências que se utiliza de técnicas estatísticas com objetivo de propor interpretações dos dados geológicos com significância estatística. A geoestatística tem se desenvolvido como importante segmento de pesquisa nas geociências de modo pioneiro como verificado em estudos de Matheron (1962, 1963 e 1965).

Diversos ramos das Geociências se valem da estatística como ferramenta já indissociável do manejo dos dados e resultados das pesquisas. Áreas como geoprocessamento de dados, prospecção mineral, entre outras, aplicam de modo consolidado os fundamentos da geoestatística. O presente estudo relaciona-se à geoestatística na medida em que emprega em sua metodologia conhecimentos da estatística para análise e suporte à estratigrafia química, fazendo assim uma conexão entre ambas as áreas, estratigrafia química e geoestatística.

A seguir serão listados de acordo com Andreotti (2003) importantes conceitos da ciência estatística para aplicação nas geociências e especificamente para auxílio na compreensão da metodologia que será empregada neste estudo.

- Estatística é a parte da ciência responsável pela coleta, organização e interpretação de dados experimentais e pela extrapolação dos resultados da amostra para a população.

- População é qualquer conjunto dos dados que constituem todo o universo de informações de que se necessita.
- Amostra corresponde a um grupo representativo da população.
- Variância é uma medida de dispersão que verifica a distância entre os valores da média aritmética.

5. Metodologia

A abordagem para extração da informação contida em bancos de dados por meio de métodos estatísticos é um dos desafios para o adequado aproveitamento do cada vez maior volume de dados.

Neste trabalho serão para tanto empregados métodos de análise estatística, sendo um deles do tipo inferencial dado pelas retas de correlação e o outro do tipo exploratório, ou seja, técnica que visa resumir as características principais para vastos conjuntos de dados de modo a frequentemente simplificar a compreensão dos dados, comumente fazendo uso de métodos visuais (Pawlowsky-Glahn, V. *et al,* 2006).

5.1 Banco de dados geoquímicos

Os Bancos de dados de modo geral são constituídos pelas mais variadas áreas de pesquisa podem atingir com o desenvolvimento das técnicas computacionais de armazenamento a casa de milhões de dados processados simultaneamente Hair, J. F. *et al.* (1995). Nesse sentido, para banco de dados

de qualquer natureza, a utilização de métodos estatísticos aponta como melhor gerenciar as quantidades expressivas de dados gerados a partir das análises químicas realizadas que por vezes, a depender da quantidade de área amostrada, podem atingir a casa de milhares de dados numéricos.

Para este estudo a dimensão do banco de dados geoquímicos de análises para o Alogrupo Guaritas atingi somente a casa das centenas de dados. Porém, ainda sim, oferece dificuldade para o desdobramento completo da informação nele contida. Os bancos de dados geoquímicos são naturalmente constituídos pelos resultados laboratoriais, que representam valores numéricos diversos para cada amostra analisado, sendo esses valores as variáveis que caracterizam cada amostra. Por isso os métodos estatísticos nesse contexto oferecem suporte adequado para o tratamento desse tipo de dado, pois as ferramentas estatísticas se aplicam justamente a grandes volumes numéricos de dados para diversas variáveis do objeto de estudo amostrado.

O problema relativo à utilização de dados geoquímicos brutos e sua interpretação foi detalhadamente descrito desde os trabalhos de Atchison (1982, 1986), o qual propôs a maneira como dados composicionais precisam ser a priori devidamente tratados antes de seu emprego direto em análises estatísticas subsequentes. Contudo, observamos a dependência natural entre os dados que compõem o objeto de estudo, o que provoca vieses em qualquer avaliação estatística de bases de dados brutos, fato que vem sendo estudado por diversos pesquisadores (Pawlowsky-Glahn & Egozcue, 2006; Daunis-I-Estadella *et al*., 2006, Melo Júnior *et al*., 2006).

Dessa forma, é parte integrante da metodologia para análise estatística o adequado manejo e compreensão do formato do banco de dados seja geoquímico ou não a ser analisado

5.2 Métodos estatísticos

As técnicas empregadas para análise estatística se deram por meio de retas de correlação entre a composição de alguns dos elementos químicos de maior interesse presentes no Alogrupo Guaritas. Também foi empregado o método estatístico de análise multivariada conhecido como Análise das Componentes Principais – ACP. Por se tratarem de duas abordagens estatísticas distintas, serão a seguir apresentadas separadamente.

5.2.1 Retas de Correlação

Esse método consiste na construção de retas de correlação por aproximação de pontos de uma ou mais variáveis pelo método dos mínimos quadrados, com cálculo da significância estatística da correlação estabelecida para os pontos amostrados, conforme Manly (1986).

Considere duas variáveis x e y que desejamos analisar se há uma correlação linear dada por y = β x + α entre elas. Tomando n amostras onde estas variáveis possam ser medidas, obtemos um conjunto de pontos $\{(x_1, y_1),(x_2, y_2),\ldots,(x_n, y_n)\}$ que podem ser colocados em um mesmo sistema de coordenadas. Neste sistema de coordenadas, podemos buscar a reta y = bx + a

que melhor aproxima a distância a estes dados pontos. Utilizando o método dos mínimos quadrados, obtemos as seguintes expressões para b e a:

$$b=\frac{n\sum x_i y_i-\sum x_i \sum y_i}{n\sum(x_i^2)-(\sum x_i)^2}, a=\bar{y}-b\bar{x},$$

onde $\bar{x}$e $\bar{y}$ são respectivamente as médias das variáveis x e y observadas nas amostras.
Espera-se que os valores amostrais de b e de a encontrados sejam boas aproximações dos valores reais β e α, respectivamente.

Assim, a **reta y = bx + a** é a chamada **reta de regressão populacional da amostragem.**

5.2.2 Significância estatística

Esse é um importante termo empregado pela estatística para estabelecer dentro de qual intervalo, ou em outras palavras, nível de confiança uma estimativa ou reta de correlação estará estabelecida com base nos dados amostrados da realidade. O pressuposto de que não houve alterações significativas entre as amostras ou nas correlações entre elas em estatística é chamado de hipótese nula, ou seja, de que não houve mudança na variável observada.

Para verificação de que pode ter havido mudanças significativas nas variáveis observadas é realizado o teste para a hipótese nula. Este teste verifica

a partir de um nível de significância pré - estabelecido se as mudanças na variável se deram acima do nível de significância estabelecido.

A partir disso, será analisado de modo formal matematicamente, conforme Aitchison, J. (1982), alguns fatores; primeiramente, que os valores de b próximos de zero indicam que a amostra não evidencia correlação linear entre x e y, dado que a influência de x na relação y = bx + a fica próxima a um termo nulo. Sendo assim, o primeiro passo na análise é o de realizar um teste estatístico de inferência para a hipótese de termos β = 0. Se a hipótese nula não puder ser rejeitada (com significância de pelo menos 95%), a amostragem não permite assegurar que haja uma correlação linear entre as variáveis x e y.

Este teste de hipótese deve ser feito utilizando uma distribuição estatística do tipo **t de Student com graus de liberdade (GL) = n-2**, i.e., com n-2 graus de liberdade. Para tanto, utiliza-se a seguinte fórmula:

$$t_{cal}=\frac{b\sqrt{\sum(x_i-\bar{x})^2}}{\sqrt{s^2}},$$

onde s^2é a variância do erro da amostragem, calculada por

$$s^2=\frac{\sum(e_i)^2}{n-2},$$

para $e_i=y_i-(b\,x_i+a)$sendo a diferença entre o valor real da amostragem y_ie o valor estimado $\hat{y}_i=b\,x_i+a$pela correlação linear indicada pela amostragem. A

hipótese nula fica rejeitada com certa significância α se $|t_{cal}|>t_{\alpha/2}$, onde $t_{\alpha/2}$é o respectivo valor da distribuição t com n-2 graus de liberdade.

Caso a hipótese β = 0 tenha sido rejeitada, encontramos um intervalo de confiança em torno do valor b para termos estimativas inferiores e superiores para o coeficiente β.

Em sequência, o próximo passo utilizado é o da medição do **coeficiente de correlação linear amostral** (também conhecido como coeficiente de correlação de Pearson), que pode ser obtido através da expressão:

$$r=\frac{\sum(x_i y_i)-n\bar{x}\bar{y}}{\sqrt{\sum(x_i)^2-n(\bar{x})^2}\sqrt{\sum(y_i)^2-n(\bar{y})^2}}.$$

Este valor r é sempre um número entre -1 e 1. Quando r > 0, dizemos que há **correlação positiva** entre x e y. Quando r < 0, dizemos que há **correlação negativa** entre as variáveis. Além disso, valores de r próximos de 1 ou de -1 indicam que a correlação é **forte**.

Por outro lado, se o valor de r estiver próximo de zero, a correlação encontrada entre as variáveis através da amostragem é **fraca**. Sendo assim, um segundo teste pode ser feito para rejeitar a hipótese ϱ = 0, onde ϱ é o coeficiente de correlação linear entre as variáveis x e y. Este teste rejeita a hipótese nula com significância α se $|t_{cal}|>t_{\alpha/2}$, onde $t_{cal}=\frac{r\sqrt{n-2}}{1-r^2}$e $t_{\alpha/2}$ é novamente o respectivo valor da **distribuição t de Student com n-2 graus de liberdade.**

Os cálculos para estabelecimento da significância estatística foram realizados manualmente, conforme a sequência de procedimentos matemáticos apresentados acima. Os gráficos com as retas de correlação estabelecidas foram produzidos por meio do software livre Geogebra[1].

5.2.3 Análise das Componentes Principais Multivariadas

A Análise de Componentes Principais (ACP), é uma das técnicas mais antigas e amplamente utilizada. Alguns exemplos de aplicação são propostos por Hair Jr. *et al.* (1995). Sua abordagem é relativamente simples - reduzir a dimensionalidade de um conjunto de dados, preservando o máximo de "variabilidade", ou seja, de informações estatísticas possível. Embora seja usada e às vezes adaptada para muitas disciplinas diferentes, trata-se de uma técnica estatística e, portanto, muito de seu desenvolvimento foi feito por estatísticos.

Assim, a ACP está associada à ideia de redução de massa de dados, com menor perda possível da informação. Contudo, é importante ter uma visão conjunta das técnicas de análise estatística multivariada de modo geral, que também associam-se à abordagem de resolução dos problemas práticos em diversas áreas do conhecimento por meio da redução de massa de dados, com menor perda possível da informação. Na ACP, especificamente, utiliza-se o procedimento matemático de transformação ortogonal para assim redistribuir a variação observada nos eixos originais de forma a se obter um conjunto de eixos ortogonais não correlacionados Manly (1986) e Hongyu (2015).

Assim, a ACP é uma técnica que consiste na transformação ortogonal de um conjunto de observações de variáveis originais possivelmente correlacionadas num conjunto de valores de variáveis linearmente não correlacionadas chamadas de componentes principais. Em outras palavras, ACP é uma técnica que transforma linearmente um conjunto original de variáveis, inicialmente correlacionadas entre si, num conjunto substancialmente menor de variáveis não correlacionadas que contém a maior parte da informação do conjunto original.

O objetivo dessa técnica é encontrar um meio de condensar a informação contida em várias variáveis originais em um conjunto menor de variáveis estatísticas, sendo estas as componentes, sempre com uma perda mínima de informação. O número de componentes principais será o número de variáveis consideradas na análise, contudo, geralmente as primeiras componentes são as mais importantes, uma vez que explicam a maior parte da variação total. As componentes principais podem ser obtidas por duas maneiras distintas: via matriz de covariância ou via matriz de correlação.

Em ambos os casos, as novas variáveis (os PCs) dependem do conjunto de dados, em vez de serem funções básicas predefinidas e, portanto, são adaptativas em sentido amplo. Os principais usos do PCA são descritivos, em vez de inferenciais. Embora, para fins inferenciais, uma distribuição normal multivariada (Gaussiana) do conjunto de dados seja geralmente assumida, o PCA como uma ferramenta descritiva não precisa de suposições distributivas e, como tal, é muito mais um método exploratório adaptativo que pode ser usado em dados numéricos de vários tipos. Na verdade, muitas adaptações da metodologia básica para diferentes tipos de dados e estruturas foram

desenvolvidas como vemos, sobretudo, em exemplos de aplicação nas ciências médicas (Fleck et al. 1998).

O método da ACP busca preservar o máximo possível de variabilidade isso significa encontrar novas variáveis que sejam funções lineares daquelas no conjunto de dados original, que maximizam a variância sucessivamente e que não estão correlacionadas entre si. Encontrar essas novas variáveis, os componentes principais (CP), consiste no estabelecimento de autovalores e autovetores, que são parâmetros estatísticos.

A obtenção desses parâmetros se dá por meio do cálculo da matriz de covariância, estabelecida a partir do banco de dados utilizado no estudo. Os autovalores obtidos representam o comprimento dos eixos das componentes principais e são medidos em unidade de variância. Associados a cada autovalor existe um vetor de módulo unitário que é o autovetor. Os autovetores representam as direções dos eixos das componentes principais.

Os procedimentos matemáticos e estatísticos necessários para o cálculo da matriz de covariância e posterior definição das CPs foram efetuados pelo software livre Jamovi[2], no qual foi inserida a base de dados geoquímicos completa.

O método das Componentes Principais para que apresente a melhor eficácia possível deve ter, pelo menos, em cinco vezes maior o número de pontos amostrados em relação ao número de variáveis observadas. Sendo assim, neste estudo para utilização desse método nos 26 pontos amostrados nas litologias do Alogrupo Guaritas optou-se por utilizar apenas os dados de 5 dos elementos químicos analisados em rocha total. Isso porque se fossem

levados em conta para a geração das CPs todos os 57 elementos químicos medidos poderia haver pouca variância para as componentes principais obtidas, uma vez que o número de amostras é de apenas 26, sendo que o número de amostras necessita ser justamente maior em ao menos 5 vezes do que o de variáveis analisadas.

Portanto, foi necessário elencar 5 elementos químicos de interesse para o estabelecimento e verificação de qual ou quais deles mais contribui para variabilidade e caracterização da composição química que distingua as sutis variações nas litologias do Alogrupo Guaritas. Para tal, foram selecionados para a Análise das Componentes Principais os seguintes compostos: Al2O3; Fe2O3; MgO; CaO e Zr.

O critério de escolha para os óxidos selecionados foi a premissa dada pelo conhecimento prévio obtido pelos estudos de estratigrafia química nos quais verifica-se que variações no teor de alumínio podem ser distintivas para a litologias estudadas. Nesse mesmo sentido foram selecionados para o estudo a quantidade total de óxido de ferro e dos óxidos de cálcio e magnésio. Dos elementos menores, optou-se por utilizar no estudo o zircônio por este ter apresentado boa correlação com outros elementos nos testes de significância estatística para as retas de correlação efetuadas.

6. Resultados obtidos

Os resultados deste trabalho encontram-se no manuscrito em anexo. Neste artigo foi proposto a aplicação de métodos estatísticos para tratamento dos dados geoquímicos disponíveis no trabalho de Denalle (2013), referentes às análises químicas de elementos maiores, menores e traço em rocha total para o Alogrupo Guaritas, Bacia do Camaquã – RS.

6.1 Carta de Submissão

17/10/2021 23:32 Chasque Webmail :: Confirming submission to Journal of South American Earth Sciences

Confirming submission to Journal of South American Earth Sciences

De	Journal of South American Earth Sciences <em@editorialmanager.com>
Remetente	<em.sames.0.76bad2.a789b6d0@editorialmanager.com>
Para	Ariane Kravczyk Bernardes <ariane.kravczyk@ufrgs.br>
Responder p...	Journal of South American Earth Sciences <support@elsevier.com>
Data	2021-10-17 23:04

This is an automated message.

Statistical methods applied to geochemical analysis of data from the Guaritas Allogroup, Camaquã Basin – RS, Brazil.

Dear Mrs. Bernardes,

We have received the above referenced manuscript you submitted to Journal of South American Earth Sciences.

To track the status of your manuscript, please log in as an author at https://www.editorialmanager.com/sames/, and navigate to the "Submissions Being Processed" folder.

Thank you for submitting your work to this journal.

Kind regards,
Journal of South American Earth Sciences

More information and support

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

#AU_SAMES#

To ensure this email reaches the intended recipient, please do not delete the above code

---

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: https://www.editorialmanager.com/sames/login.asp?a=r). Please contact the publication office if you have any questions.

6.2 Manuscrito

Statistical methods applied to geochemical analysis of data from the Guaritas Allogroup, Camaquã Basin – RS, Brazil.

ARIANE KRAVCZYK BERNARDES & ANA MARIA PIMENTEL MIZUSAKI
Postgraduate Program in Geosciences, Institute of Geosciences, Federal University of Rio Grande do Sul, Av. Bento Gonçalves, 9500, Cx.P. 15001, 91501-970, Porto Alegre, RS, Brazil. ariane.kravczyk@ufrgs.br, ana.mizusaki@ufrgs.br

ABSTRACT: In this work, we use statistical methods for the analysis of geochemical data obtained through the technique of sample fusion mass spectrometry with inductively coupled plasma (FUS-ICP: Fusion-Inductively Coupled Plasma). The statistical treatment of geochemical data provides clusters and correlations of major, minor and trace chemical elements present in the chemical composition of the rocks of the Guaritas Allogroup, belonging to the Camaquã sedimentary basin, located in southern Brazil, in the state of Rio Grande do Sul. In addition to contributing to the consolidation of knowledge about the Camaquã Basin - RS/Brazil, the focus of this study is the statistical analysis of geochemical data available for the Guaritas Allogroup, from the work of Denalle (2013). The following statistical methods were applied: straight correlation by approximating the points of one or more variables by the method of least squares, with calculation of the statistical significance of the correlation established for the sampled points and Principal Component Analysis – PCA. The statistical approach aims to optimize the data available for the Guaritas Allogroup, with the purpose to refining the information and possible interpretations of the data, which better establish the compositional characteristics and genesis relationships for these rocks. At the sampled points there are subtle variations in chemical composition, but no significant macroscopic variations were observed. The establishment of correlation lines helps to understand the specific relations of the composition of a chemical element, which can be a major element in the rock, for example, with respect to the variation of one or two minor elements or traces. The Principal Component Analysis method – PCA contributes to the understanding of the influence of which sets of chemical elements are decisive in the differentiation of lithological types. This technique proves to be advantageous for the analysis of extensive databases, as it reduces the excessive amount of data, eliminating overlaps, but with little loss of significant information. PCA points to choices of the most representative forms of data through linear combinations of the original variables. This technique provides advantages in the management of geochemical data or other areas of research in geosciences, and can be used in an analogous way in several available databases. For the Guaritas Allogroup, this technique proved to be useful in pointing out compositional relationships between major elements, main rock formers, such as SI, Ca, Al and Mg, and minor elements, such as Mn, Zn, Ti, Rb and Sr. Therefore, the preliminary results obtained represent an advantageous tool, used in an additional way, for the interpretation of geochemical data in support of the analytical techniques traditionally employed in chemical stratigraphy.

Keywords: Statistical Analysis, Principal Components, Chemical Stratigraphy, Ordovician, Allostratigraphy.

1. Introduction

Geochemistry is an applied analytical tool that helps to understand geological processes such as recognition of paleoenvironments, paleoclimates, tectonics, provenance and stratigraphic correlations in sedimentary basins, among other processes. However, the characterization of depositional units with a focus on distinction and correlation based on the variation of chemical data and the use of the term “chemostratigraphy” or “chemical stratigraphy” for this were more frequently used from the 1980s onwards, as identified in Ramkumar (2015).

Chemostratigraphy covers the identification of geochemical characteristics based on the spatiotemporal variation of the chemical elements present in rocks or even on the variation of the ratios between these elements (Ramkumar, 2015 apud Prundeanu, 2021). Thus, chemostratigraphy has been applied in studies of a large number of basins and sedimentary units around the world, such as the Foreland Andean Basins (Pinto et al, 2018), Karoo Basin, South Africa (Franchi el al, 2021) and Basin of Duero in Spain (Gómez et al, 2006). However, challenges arise regarding how to better manage the extensive amount of data generated from the chemical analyzes carried out that sometimes, depending on the amount of sampled area, can reach thousands of numerical data points.

Thus, despite the problem related to the use of raw geochemical data and its interpretation have already been described in detail since the works of Aitchison (1982, 1986), who proposed how compositional data needs to be properly treated a priori before its direct use in subsequent statistical analyses. Even so, we observe the natural dependence between the data that make up the object of study, which causes biases in any statistical evaluation of raw databases, a fact that has been studied by several researchers (Pawlowsky-Glahn & Egozcue, 2006; Daunis- I-Estadella et al., 2006, Melo Júnior et al., 2006).

Thus, we verified the relevance of the set of statistical techniques used in geoscience research, constituting the bases for acting in the field of Geostatistics. This segment has developed as an important area of research in geosciences in a pioneering way as verified in studies by Matheron (1962, 1963 and 1965).

Over the decades since the mid-1960s, studies have shown the effectiveness of using Geostatistics as an alternative analytical tool among others already established in research in natural sciences, Andriotti (2003). Therefore, the term Geostatistics is designated to represent the specific study area in geosciences that uses statistical techniques in order to propose interpretations of geological data with statistical significance.

In this work, statistical methods will be used from the analysis of geochemical data obtained from the stratigraphic unit called Guaritas Allogroup, located in the Camaquã Basin in the state of Rio Grande do Sul, Brazil in which, first, the application of chemostratigraphy was proposed by Denalle (2013).

The samples analyzed in this basin correspond to the portion of the Guaritas Allogroup, which all have similar lithological types. Thus, the variation in the classification of sedimentary rocks observed for this region is more limited with similar granulometric and structural criteria. While there is greater variation in the chemical composition of the samples studied at the expense of the lithological types. Therefore, we seek to detail the differences present in the studied samples, albeit subtle, in their chemical composition, especially with regard to the relationship between the presence of major elements and the occurrence of certain minor elements.

The main proposition of this work will be to apply statistical techniques to refine the geochemical data obtained for the Guaritas Allogroup (Ordovician of the Camaquã Basin), proposed by Denalle (2013) in order to obtain relationships with statistical significance, which can serve as a basis for recognition of patterns of chemical

elements, present in mineral composition, capable of providing geological interpretations of genetic correlations and, in certain cases, of provenance in sedimentary basins.

The statistical approach will be through two main sets of methods, namely: Correlation line by approximation of points of one or more variables by the method of least squares, with calculation of the statistical significance of the correlation established for the sampled points; Principal Component Analysis – PCA, a method in which the mathematical procedure of orthogonal transformation of a set of observations of variables, possibly correlated, is carried out in a set of values of linearly uncorrelated variables, called principal components.

1.1 Study Area

The Camaquã Basin, hereinafter represented by the acronym “CB”, is characterized by a context of geological complexity, as its stratigraphic record presents a wide variety of lithological types. In CB there are from sedimentary rocks of marine, alluvial and wind origin to volcanic and volcaniclastic rocks of varied compositions. Intense successive tectonic events were responsible for the CB deformation, among which the deposition of sedimentary rocks occurred. This scenario contributed to a complex geological environment, Paim et al (2000).

The first studies of CB date back to the early 1930s (eg Carvalho, 1932). Since then, numerous works have been carried out over the following decades providing an abundance of interpretations and models. From the final decades of the 20th century, with an approach predominantly influenced by the theory of Plate Tectonics, research culminated in a synthesis of the evolution of the Camaquã Basin from the perspective of tectono-sedimentary cycles superimposed on the same depositional locus (Paim et al, 2002).

However, the groups and formations belonging to CB still offer real challenges in the search for an understanding of its genesis and paleoenvironmental context, especially regarding the geochemistry and geochronology of sedimentary rock formations. It is proposed the division of CB into allogroups Maricá, Bom Jardim, Cerro do Bugio, Santa Bárbara and Guaritas (Borba, 2006), with the latter groups, Santa Bárbara and Guaritas representing the last units of the Camaquã Basin, formed in a tectonic environment predominantly transtrational, with depositional age ranges still being controversial (Oliveira, 2012).

The Guaritas Allogroup is located in the central region of the state of Rio Grande do Sul, on the Rio-Grandense Sul Shield, between the parallels -54°00'00" and -53°00'00" of latitude and the meridians -30 °00'00" and -31°00'00" in longitude, south of the municipality of Caçapava do Sul and west-northwest of the municipality of Santana da Boa Vista, extending north to the region of Encruzilhada do Sul, and about 250 km from the city of Porto Alegre by the BR-290 road.

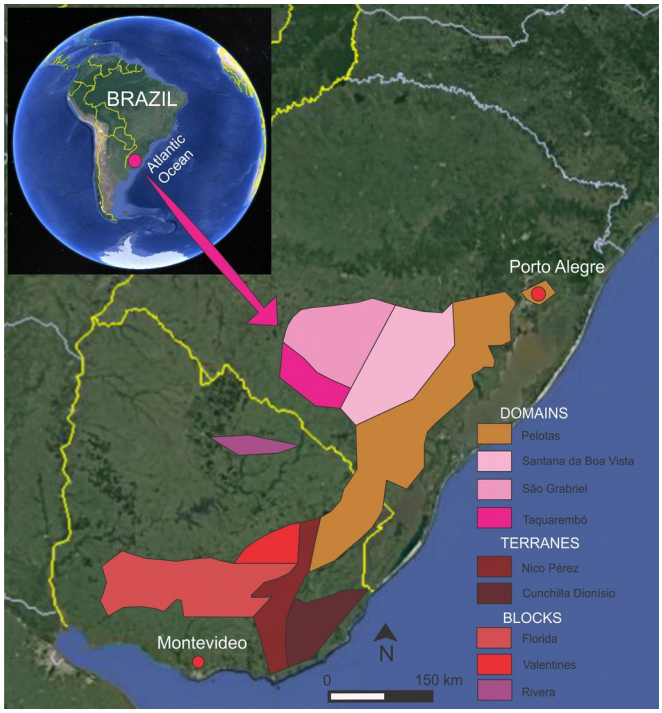


Figure 1 – Location and geological context of the Guaritas Allogroup, Camaquã Basin, Rio Grande do Su – Brazil.

The Guaritas Allogroup is considered to be the last stage of deposition in CB, it outcrops in an elongated band in the central region of the Rio-Grandense Sul Shield, and t is formed by two large sequences: Pedra Pintada (base) and Varzinha (top). Radiometric (U/Pb) zircon dating from the volcanic base of the Pedra Pintada sequence suggests an age of 470 ± 19 Ma for the magmatic event, therefore having a probable Ordovician age for the sequence (Maraschin et al., 2010).

In this study, the nomenclature used for the CB lithological units will be given by allostratigraphy. This classification considers the stratigraphic units delimited by discontinuity surfaces that involve these units and that allow their geological mapping. Since, while in lithostratigraphy the main criterion is lithological identity or association, in allostratigraphy there may be rocks from more than one type of sedimentary environment, therefore, lithologically differentiated, but marked by non-deposition or erosion surfaces that correlate them both laterally and temporally.

This perspective given by allostratigraphy is appropriately consistent with the choice of chemical stratigraphy used to study and characterize the lithologies of the Guaritas Allogroup and their alloformations and allomembers, and the distinction of the chemical composition of major, minor and trace elements can establish genetic relationships between apparently different lithologies. Thus, statistical analysis serves as a tool capable of qualitatively improving the data provided by chemical stratigraphy.

The area selected for sample collection is comprised of a columnar profile which is around 48 meters thick from the base to the top, and 26 rock samples were collected along this columnar profile at previously established points. Sampling was carried out in a precise and representative way to carry out laboratory analyzes with adequate control of the stratigraphic positioning of the samples along the columnar profile (Denalle. 2013).

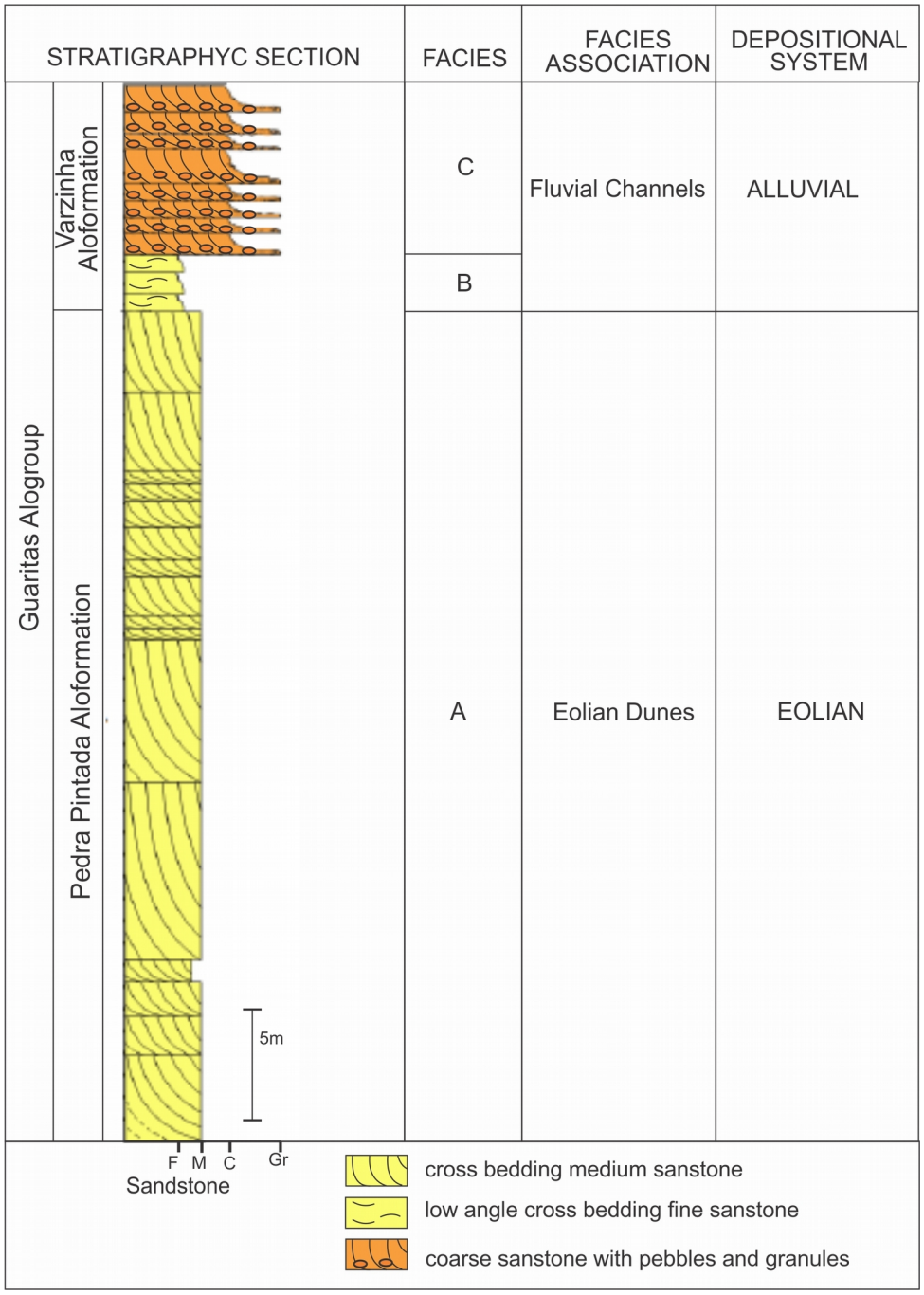


Figure 2 -Stratigraphic column for the Guaritas Allogroup, corresponding to the region sampled for generating the database (modif. Denalle, 2013).

The lithologies corresponding to the sampled points consist macroscopically of reddish-brown sandstones of alluvial, lacustrine, fluvial and eolian origin. These lithologies are inserted in the final filling phase of the Camaquã Basin, with ages of 470 ± 19 Ma (Ordovician) for the volcanic rocks at the base of the Guaritas Allogroup, obtained by dating a zircon grain by the U/PB method, Hartmann et al. (1998). Dates were also carried out using the K/Ar method on autogenic illites contained in strata located just above the volcanic rocks at the base of the Pedra Pintada Alloformation, proposed by Maraschin et al. (2010), indicating ages between 473.7 ± 9.4 and 521.7 ± 10.3 Ma (Lower Cambrian to Ordovician).

1.2 Objectives

The demand for the use of analytical databases, supplied on the scale of thousands of numerical values, is increasingly present in many areas of knowledge, for instance, in the biological and health sciences. From the accurate analysis of the data, it is possible to generate interpretations based on a large volume of data. In this sense, several statistical methods of multivariate analysis have been developed with an emphasis on managing a large amount of available data.

One of the advantages of certain sets of statistical techniques is that they considerably reduce the dimensionality of large databases. However, this reduction mechanism aims to maintain an interpretable form of information, so that most of it is preserved in the databases. Many statistical techniques have been developed for this purpose. Principal Component Analysis (PCA), however, is one of the oldest and most widely used techniques. Some application examples are proposed by Hair Jr. et al. (1995). His approach is simple - to reduce the dimensionality of a dataset while preserving as much "variability", i.e., statistical information as possible. Although PCA is used and sometimes adapted for many different disciplines, it is a statistical technique and therefore much of its development has been done by statisticians.

The use of the correlation lines and the PCA will be used in order to fulfill the main objective of this study, which is the refinement of the data available for the Guaritas Allogroup. Thus, the effectiveness of employability of the techniques that will be tested for the set of geochemical data available for that region will be verified.

2. Methodology

In this work, the aforementioned database from the work of Denalle (2013) was used. The database consists of 26 sampled points in which analysis were performed for 57 chemical elements, including major, minor and trace elements. The analytical technique used to detect the elements was mass spectrometry by fusion of the sample with inductively coupled plasma (FUS-ICP: Fusion-Inductively Coupled Plasma). Whole rock chemical analysis for major, minor and trace elements were prepared at Activation Laboratories Ltd. (Actlabs Laboratory) based in Canada.

The techniques used for statistical analysis were through correlation lines between the composition of some of the most interesting chemical elements present in the Guaritas Allogroup. The statistical method of multivariate analysis known as Principal Component Analysis – PCA was also used. As they are two different statistical approaches, they will be presented below in separate sections, one for each technique.

2.1 Correlation Lines

This method consists of the construction of correlation lines by approximating the points of one or more variables by the method of least squares, with calculation of the statistical significance of the correlation established for the sampled points. Consider two variables x and y that we want to analyze wether there is a linear correlation given by $y = \beta x + \alpha$

between them. Taking n samples where these variables can be measured, we obtain a set of points $\{(x_1, y_1),(x_2, y_2),\ldots,(x_n, y_n)\}$ that can be plotted in the same coordinate system. In this coordinate system, we can find the line y = bx + a that best approximates the distance to these given points. Using the least squares method, we obtain the following expressions for b and a:

$$b=\frac{n\sum x_i y_i-\sum x_i \sum y_i}{n\sum(x_i^2)-(\sum x_i)^2}, a=\bar{y}-b\bar{x},$$

where $\bar{y}$ and $\bar{x}$ are respectively the means of the variables y and x observed in the samples.

It is expected that the values of b and a found are good approximations for the actual values β and α, respectively.

Thus, the line y = bx + a is called the sampling population regression line. From this, we must analyze some factors. First, that values of b close to zero indicate that the sample does not show a linear correlation between x and y, given that the influence of x on the relationship y = bx + a is close to a null term. Therefore, the first step in the analysis is to perform a statistical inference test for the hypothesis β = 0. If the null hypothesis cannot be rejected (with a significance level of at least 95%), sampling does not ensure that there is a linear correlation between the x and y variables.

This hypothesis test should be done using a Student's t-type statistical distribution with (GL) = n-2 degrees of freedom, i.e., with n-2 degrees of freedom. To do so, the following formula is used:

$$t_{cal}=\frac{b\sqrt{\sum(x_i-\bar{x})^2}}{\sqrt{s^2}},$$

where s^2 is the variance of the sampling error, calculated by

$$s^2 = \frac{\sum (e_i)^2}{n-2},$$

for $e_i = y_i - (b\,x_i + a)$ being the difference between the actual value of the sample y_i and the estimated value $\hat{y}_i = b\,x_i + a$ by the linear correlation indicated by sampling. The null hypothesis is rejected with some significance α if $|t_{cal}| > t_{\alpha/2}$, where $t_{\alpha/2}$ is the respective value of the t distribution with n-2 degrees of freedom.

If the hypothesis β = 0 has been rejected, we find a confidence interval around the value b to have lower and higher estimates for the coefficient β.

The next step used is the measurement of the sample linear correlation coefficient (also known as Pearson's correlation coefficient), which can be obtained through the expression:

$$r = \frac{\sum (x_i y_i) - n\,\bar{x}\,\bar{y}}{\sqrt{\sum (x_i)^2 - n(\bar{x})^2}\sqrt{\sum (y_i)^2 - n(\bar{y})^2}}.$$

This r value is always a number between -1 and 1. When r > 0, we say that there is a positive correlation between x and y. When r < 0, we say that there is a negative correlation between the variables. Also, values of r close to 1 or -1 indicate that the correlation is strong. On the other hand, if the value of r is close to zero, the correlation found between the variables through sampling is weak. Thus, a second test can be done to reject the hypothesis ϱ = 0, where ϱ is the linear correlation coefficient between the variables x and y. This test rejects the null hypothesis with significance α if $|t_{cal}| > t_{\alpha/2}$,

where $t_{cal} = \frac{r\sqrt{n-2}}{1-r^2}$ and $t_{\alpha/2}$ is again the respective value of the Student's t distribution with n-2 degrees of freedom.

Calculations to establish statistical significance were performed manually, according to the sequence of mathematical procedures presented above. The graphs with the established correlation lines were produced using the open source Geogebra software.

.2.2 Analysis of Main Components - PCA

PCA is associated with the idea of reducing the mass of data, with the least possible loss of information. However, it is important to have a joint view of general multivariate statistical analysis techniques, which are also associated with the approach to solving practical problems in various areas of knowledge through the reduction of data mass, with the least possible loss of information. In PCA, specifically, the mathematical procedure of orthogonal transformation is used to redistribute the variation observed in the original axes in order to obtain a set of uncorrelated orthogonal axes Manly (1986) and Hongyu (2015).

Thus, PCA is a technique that consists of the orthogonal transformation of a set of observations of possibly correlated original variables into a set of values of linearly uncorrelated variables called principal components. In other words, PCA is a technique that linearly transforms an original set of variables, initially correlated with each other, into a substantially smaller set of uncorrelated variables that contain most of the information present in the original set.

The objective of this technique is to find a way to condense the information contained in several original variables into a smaller set of statistical variables, these being the components, always with a minimal loss of information. The number of principal components will be the number of variables considered in the analysis, however, generally the first components are the most important as they explain most of the total variation. The principal components can be obtained in two different ways: via the covariance matrix or via the correlation matrix.

In either case, the new variables (the PCs) depend on the dataset rather than being predefined basic xfunctions and therefore are adaptive in a broad sense. The main uses of PCA are descriptive rather than inferential. Although, for inferential purposes, a multivariate (Gaussian) normal distribution of the dataset is generally assumed, PCA as a descriptive tool does not need distributive assumptions and, as such, is an adaptive exploratory method that can be used on numerical data of various types. In fact, many adaptations of the basic methodology for different types of data and structures have been developed as we see, above all, in examples of application in the medical sciences (Fleck et al. 1998).

The PCA method seeks to preserve as much variability as possible. This means finding new variables that are linear functions of those in the original dataset that successively maximize the variance and that are not correlated with each other. Finding these new variables, the Principal Components (PC), consists of establishing eigenvalues and eigenvectors which are statistical parameters. These parameters are obtained by calculating the covariance matrix, established from the database used in the study. The eigenvalues obtained represent the length of the main component axes and are measured in units of variance. Associated with each eigenvalue there is a unitary modulus vector which is the eigenvector. The eigenvectors represent the directions of the axes of the principal components.

The mathematical and statistical procedures necessary for the calculation of the covariance matrix and subsequent definition of the PCs were carried out by the open software Jamovi[1], in which the complete geochemical database was inserted.

In order to present the best possible efficacy, the Principal Components method must have at least five times the number of points sampled in relation to the number of variables observed. Therefore, in this study to use this method in the 26 points sampled in the lithologies of the Guaritas Allogroup, it was decided to use only data from 5 of the chemical elements analyzed in whole rock. This is because if all 57 chemical elements

measured were taken into account for the generation of the PCs, there could be little variance for the main components obtained, since the number of samples is only 26, and the number of samples needs to be at least 5 times that of the analyzed variables.

Therefore, it was necessary to list 5 chemical elements of interest for the establishment and verification of which of them contributes most to the variability and characterization of the chemical composition that distinguishes the subtle variations in the lithologies of the Guaritas Allogroup. To this end, the following compounds were selected for the Principal Component Analysis: Al2O3; Fe2O3; MgO; CaO and Zr.

The choice criterion for the selected oxides was the premise given by the previous knowledge obtained by chemical stratigraphy studies in which it is verified that variations in the aluminum content can be distinctive for the studied lithologies. In the same sense, the total amount of iron oxide and calcium and magnesium oxides were selected for the study. Of the minor elements, we chose to use zirconium in the study because it showed a good correlation with other elements in the statistical significance tests for the correlation lines performed.

3. Results

Next, the graphs developed for understanding the correlations between the chemical elements will be presented, first, in order to understand how each one individually behaves in relation to the other. For the elaboration of these graphs, the most abundant oxides measured in the samples were used and for which certain types of correlation are expected according to the chemical composition of the studied rocks. Thus, the correlation lines can serve as a statistical mean of validating the expected compositional relationships. After the presentation of the correlation graphs, the main components will be shown, ordered according to their degree of variance from largest to smallest, that is,

identifying among the 5 oxides used for PCA which were more determinant in the variability between the lithologies in the 26 points sampled.

3.1 Correlation Lines

Altogether, 18 correlations between major, minor and trait elements were proposed, 10 of them, that is, a little more than half presented some kind of correlation between the elements with statistical significance in a confidence interval of 95%. For more details on the parameters calculated for the correlations shown in each graph, see the appendix at the end of this article. The other correlations for a confidence interval arbitrated in this study of 95% showed no correlation. However, some of them presented weak correlation and for an arbitrated interval of 90% they would be considered valid correlations, being the case of the correlations between the oxides of Al2O3 and MnO; Al2O3 and Fe2O3; Fe2O3 and CaO; SiO2 and K2O and Ba and Rb.

The 95% confidence interval was arbitrated in this study to ensure that the proposed correlations are concisely established. However, all correlations are presented, even those not valid for this established statistical criterion, as they also present some statistical significance for lower confidence intervals, such as 90%. Thus, this type of information may indicate that stronger correlations can be established depending on a greater number of sampled points and analyzed rocks.

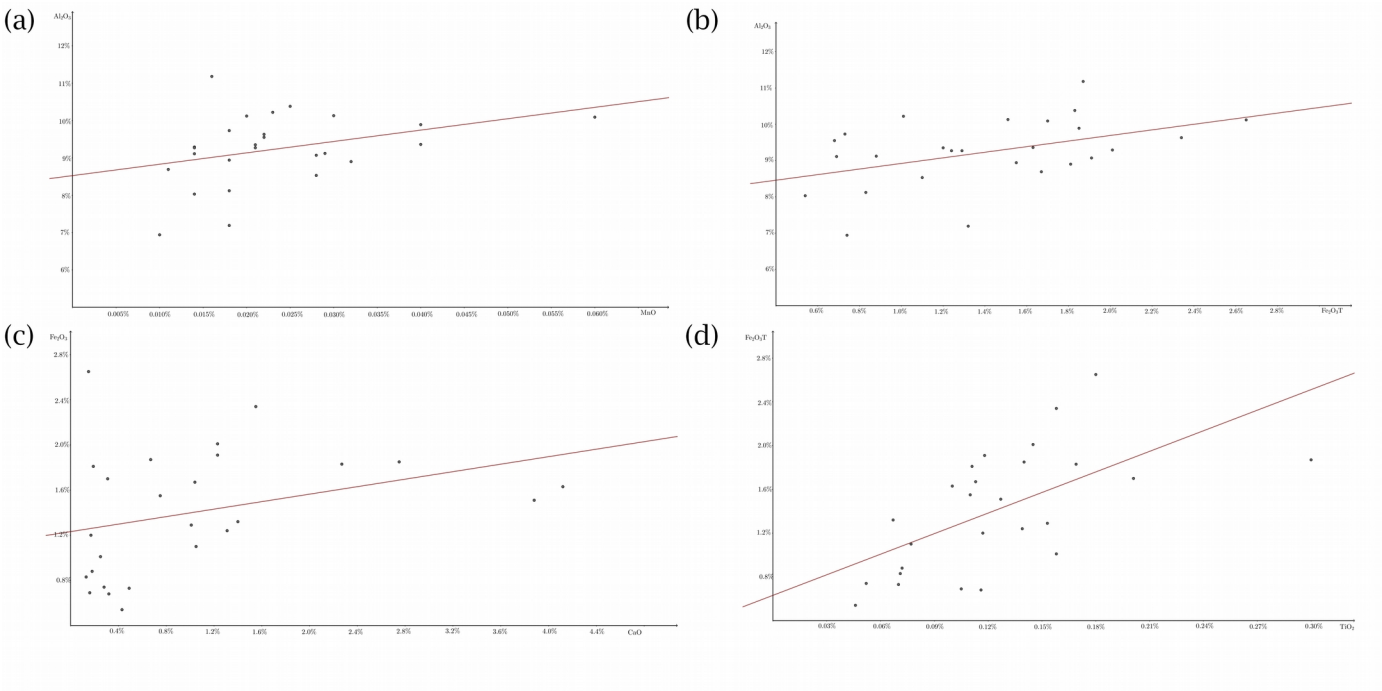


Figure 3 – Correlations for Al and Fe oxides; (a) weak correlation, that is, without statistical significance for the presence of Al oxide in relation to Mn, (b) positive correlation within the 95% confidence interval for the presence of Al oxide, which increases in relation to presence of Fe oxide, (c) correlation not found for Fe oxide in relation to the presence of Ca oxide, (d) positive correlation within the 95% confidence interval for Fe oxide which increases in relation to the presence of Ti oxide.

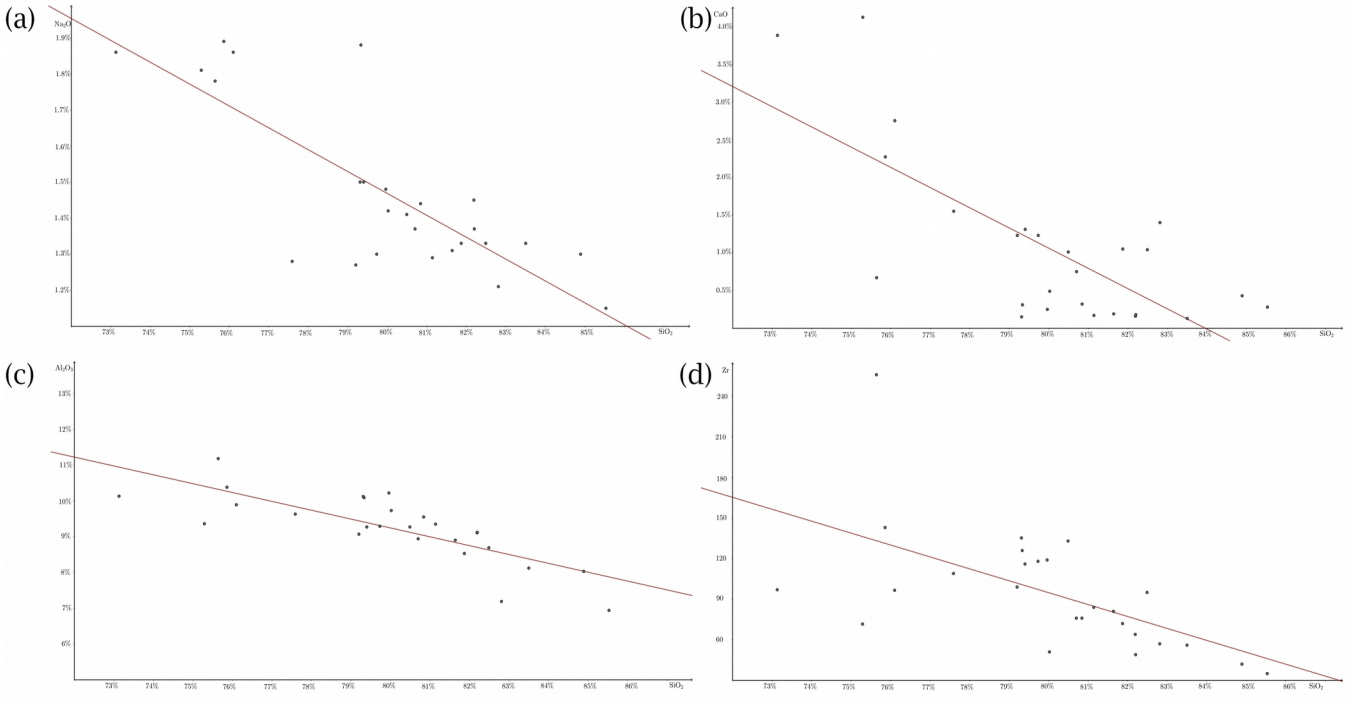


Figure 4 – Correlations for Si oxide; (a) negative correlation between Si oxide, which decreases in quantity as that of Na oxide increases, (b) negative correlation between Si oxide, which decreases in quantity as that of Ca oxide increases, (c) negative correlation between Si oxide, which decreases in quantity as that of Al oxide increases, (d) negative correlation between Si oxide, which decreases in quantity as that of the Zr element increases.

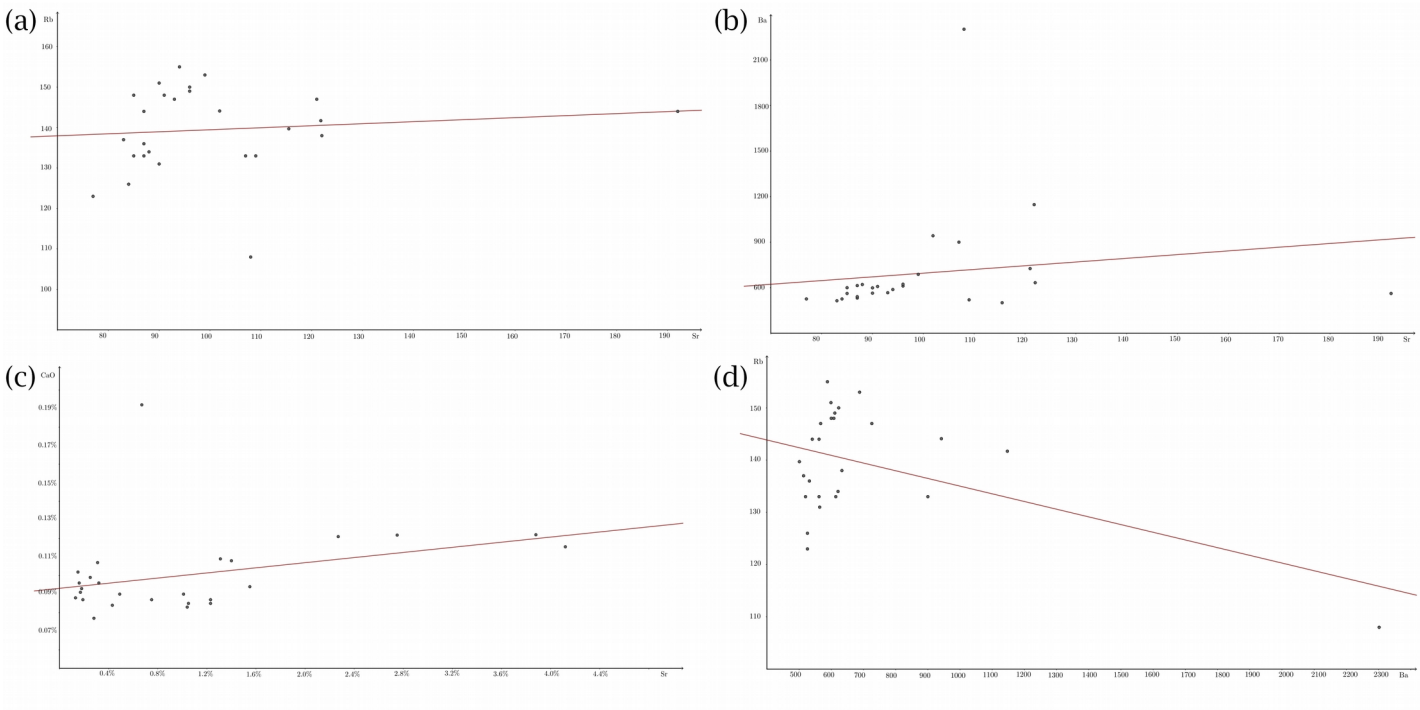


Figure 5 – Invalid correlations for the established 95% confidence interval. (a) No correlation for the minor elements Sr and Rb, (b) No correlation for the minor elements Sr and Ba, (c) No correlation for the minor element Sr with respect to Ca oxide, (d) No correlation for the minor elements Ba and Rb.

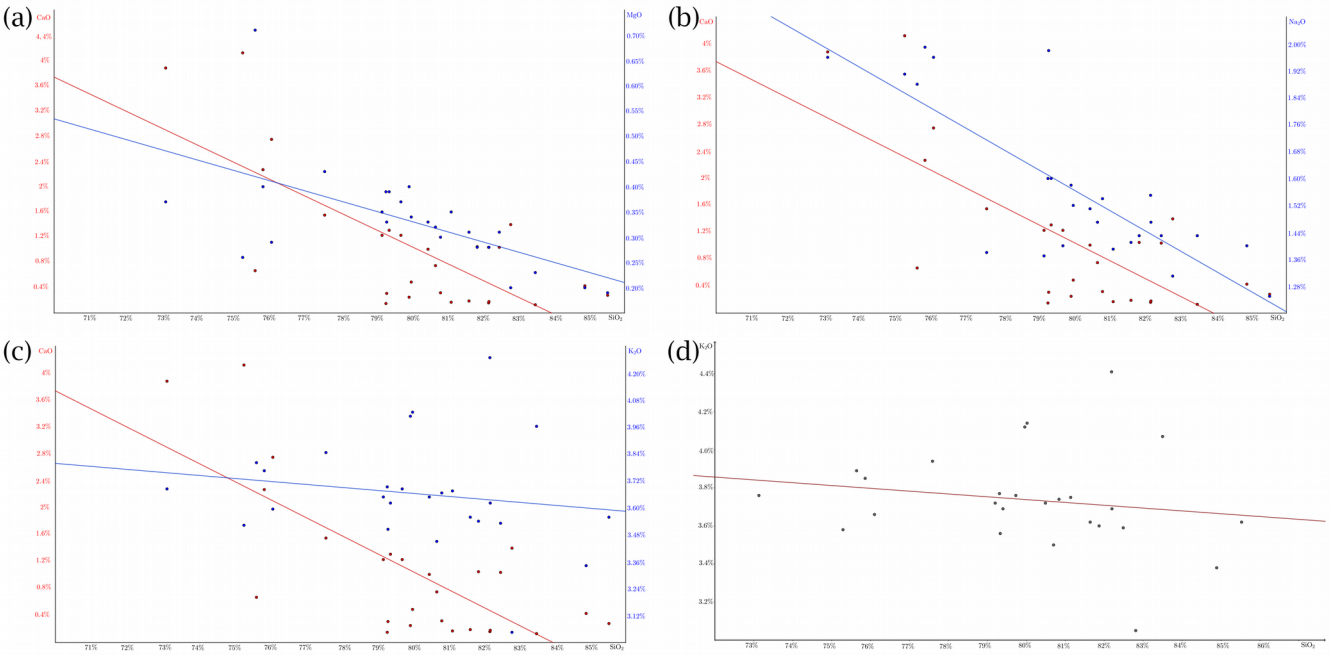


Figure 6 – Correlations between the oxides of Si, Ca, Na and K. (a) Si oxide has a negative correlation with the oxides of Ca and Mg, although there is no correlation between the oxides of Ca and Mg, (b) Oxide of Si is negatively correlated with Ca and Na oxides, (c) Si oxide is negatively correlated with Ca oxides, but does not show a valid correlation for K oxide, (d) Si oxide is not correlated with K oxide.

3.2 Main Component Analysis

After loading the database in the open software Jamovi [2], the values of the 5 variables listed for application of the main components were processed, with 4 oxides (Al2O3, Fe2O3, MgO, CaO) representing the largest elements present in the 26 points sampled in the along the columnar profile, about 38 meters thick, of the Guaritas Allogroup. Finally, a minor element, zirconium (Zr), was selected because it presented a good linear correlation with the oxides of the major elements.

To ensure the effectiveness and significance of the Principal Components method, two tests were applied to ensure its validity to determine an original set of variables, initially correlated in a smaller set of uncorrelated variables that contain most of the information from the original set.

The tests performed were Bartlett's sphericity test to assess whether there is a correlation between the original variables. To be meaningful, the test result must be less than 0.05. Also, we used the sample adequacy measure test, Kaiser - Meyer - Olkin (KMO) test. This test evaluates the proportion of common variance between the observed variables, its result must be greater than 0.5 to be considered a good sampling adequacy. It is noteworthy that the variance corresponds to a measure of statistical dispersion, that is, how much a variable can assume different values from its expected mean value. This dispersion measure applied to geochemical data can help to indicate which elements are predominantly varying along the columnar profile, and may become target elements in chemostratigraphic correlation studies.

Both Bartlett and KMO sphericity tests gave adequate results (Tables 1 and 2). Thus, the analysis proceeded, obtaining as a result 5 main components, numbered from 1 to 5, with number 1 being the component that presents the greatest variance, component number 2, the second largest variance, and so on. Note that in this statistical method, the same number of principal components as the number of original variables are generated.

Note that the principal components obtained are not the same as the original variables for each sample. The main components are the product generated by the orthogonal transformation (in this case of the "varimax" type) in which each main component generated is independent and not correlated with each other, presenting a corresponding eigenvalue and a proportion of variance that defines how each explains the variance in the samples according to the original variables (Table 3).

Bartlett's Test of Sphericity		
χ^2	**df**	**p**
77.8	10	< .001

Table 1 – Bartlett test result

KMO Measure of Sampling Adequacy	
	MSA
Overall	**0.717**
Al2O3	0.825
Fe2O3	0.794
MgO	0.666
CaO	0.409
Zr	0.709

Table 2 – KMO test result for sampling and specific for each variable.

Initial Eigenvalues

Component	Eigenvalue	% of Variance	Cumulative %
1	3.0655	61.31	61.3
2	1.0839	21.68	83.0
3	0.5412	10.82	93.8
4	0.2347	4.69	98.5
5	0.0747	1.49	100.0

Table 3 – Initial eigenvalues for each main component generated. Components 1 and 2 explain most of the observed variance, together accounting for 83% of the total variance observed for the variables of the 26 sampled points of rocks of the Guaritas Allogroup.

Among the 5 main components generated in PCA, components 1 and 2 define most of the variability observed in the original variables. We highlight component 1, which describes 61.31% of the total variance. In Table 4, we see in the middle column the values that describe how much the main component was influenced by the original variables, and this influence may even be non-existent, as in the case of the CaO variable. In the right column, we have the uniqueness information for each original variable, which is the proportion of the unique variance for each original variable, which is not given by the principal components.

Component Loadings

	Component	
	1	**Uniqueness**
Al_2O_3	0.873	0.238
Fe_2O_3	0.716	0.487
MgO	0.923	0.148
CaO		0.935
Zr	0.935	0.126

Note. 'varimax' rotation was used[3]

Table 4 – Principal component number 1 is responsible for 61.31% of the total variance.

Scree Plot

The scree plot is a widely used method to define which main components will be taken into account in exploratory studies on a given topic. Figure 7 illustrates the scree plot in which the variances of the principal components are plotted, that is, their eigenvalues in decreasing order of magnitude.

To select which components will be used in the study, this graph can be especially useful when large numbers of main components are generated in PCA and a clear selection criterion is needed for which PCs are more representative of the analyzed variables. Thus, the literature suggests selecting the principal components that are located prior to the inflection point between the data and simulation curves, at which point there is a significant drop in the explained variance in relation to the previous component. By convention, it is also common to use principal components that have an eigenvalue greater than 1.

In this study, it is verified that, for any of these adopted criteria, the main components to be used must be numbers 1 and 2, as illustrated in the scree plot graph in the figure. 7.

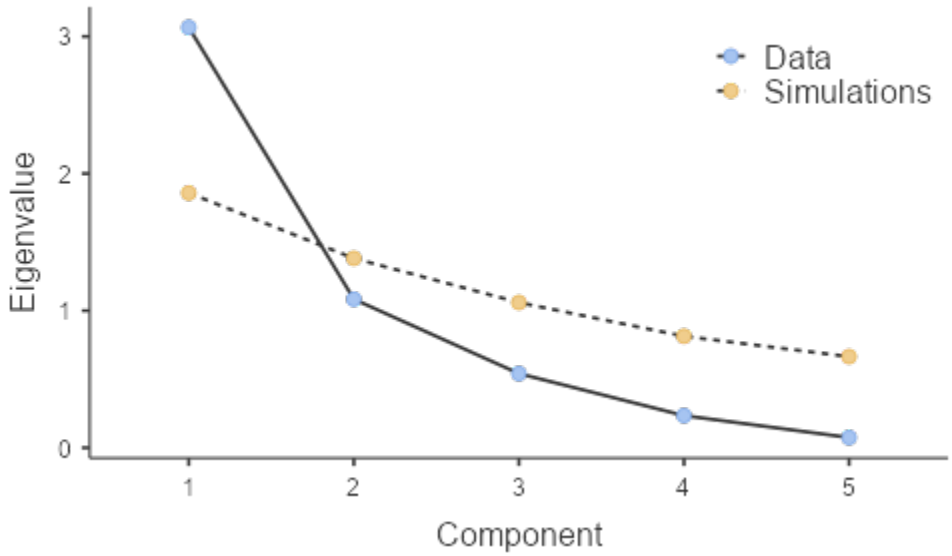


Figure 7 – Scree plot method for extracting the principal components.

4. Discussion

The results obtained with the correlation lines were partially satisfactory, since it was not possible to obtain strong correlations for all the analyzed compositions for the 95% confidence interval. However, for half of the proposed correlations there was a correlation with statistical significance. Based on this data set, it was possible to verify which oxides maintained correlations with the greatest number of elements. In this sense, silicon oxide, as expected, maintained correlations with other oxides of major elements such as Ca, Na, Mg and Al, and the only oxide with no correlation was K. Regarding minor and trace elements, it was not verified correlations between the Sr, Rb, Ba elements. Possible expected correlations between Fe2O3 and CaO, Sr and CaO oxides were also not verified. However, part of this may be associated with the fact that for statistical analysis, especially for chemical elements of lesser occurrence, such as minor and trace elements, the required sample universe needs to be larger than usually used in research, due to the high cost of chemical analysis in whole rock. The correlations verified, however, served as an aid even to identify which chemical

compositions would be more suitable for using the statistical method of analysis of the main components, as was the case of the positive correlations found between Fe2O3 and TiO2 and for Fe2O3 and Al2O3 and the negative correlation between SiO2 and Zr.

The oxides of Al2O3, Fe2O3, MgO, CaO and the element Zr were selected as the variables for the 26 sampled rock points to be analyzed by PCA. However, the other elements contained in the geochemical database for the Guaritas Allogroup, which includes about 52 analyzed elements, are also relevant variables for the 26 points where the rocks were collected for the analysis. However, for PCA to be properly employed as a statistical method, it is necessary that the number of variables does not exceed the number of samples, and more than that, it is recommended that the number of samples be at least 5 times greater than the number of variables present in the analysis. For this reason, previously mentioned in the methodology section, it was not possible to extend the PCA method to all of the more than 50 elements of the Guaritas Allogroup database.

In this way, additional statistical analyzes can be carried out in a broader way based on the performance of new chemical analyzes for this study region, thus increasing the number of samples for applying the PCA method. However, the adaptation of the technique to a smaller number of samples proved successful, as from the determination by the correlation criteria found for certain elements, added to the knowledge of which oxides most contribute to the compositional variability of the lithologies, it was possible to select 5 compositional variables only. Thus, the PCA method was statistically adequate with good results in both the Bartlett and the KMO tests.

Among the 5 variables used for PCA, the main component generated that most explains the total variance of the sample universe was strongly influenced by the variables represented by the oxides of Mg, Al and Fe, with emphasis on the influence of Zr. Only the Ca oxide did not provide sufficient influence on its proportion of variance for the main component 1. Based on the geochemical data for the Guaritas Allogroup and by the PCA method, Mg oxide and Zr were the variables, among the 5 used in the PCA, which

further define the variability for rock samples along the columnar profile in which they were collected.

Although the method only informs about which chemical compositions vary most along the columnar profile, that is, the compositions that do not appear homogeneously along the profile and the method does not identify which depth levels in the columnar profile where this compositional variation is more intense. Still, this pattern of compositional variation obtained statistically can be useful in chemostratigraphic correlation studies for which statistical methods are also applied, where similar statistical patterns of compositional variation can be verified, which can guide further field and petrographic studies and even chemical analyzes for establishing stratigraphic correlations.

5. Conclusion

The use of statistical methods proves to be advantageous in chemical stratigraphy studies, especially when data is already available. Even for an area where there is still no analysis, the statistical treatment of data from new analysis brings benefits, while it can restrict the number of target elements of the study, reducing the need for successive, more comprehensive chemical analysis, involving more than 50 distinct chemical elements. Of course, this all depends on the research objectives and interest in certain minerals and specific chemical compositions.

The establishment of correlation between the individual chemical elements was important for case-by-case verification of their proportions during the sampling carried out for the columnar profile of the Guaritas Allogroup. With the PCA method, it was possible to establish which oxides contributed the most to the variability in the composition of the sampled lithologies. Thus, for future chemostratigraphic correlation research in this area of study it may be possible to carry out chemical analysis of whole rock for a smaller number of minor and trace elements. Therefore, with PCA it was

possible to establish with statistical significance which chemical compositions are more distinctive for the regions sampled along the columnar profile in the Guaritas Allogroup.

In addition to the advantage of a possible cost reduction with chemical analysis, statistical methods are able to improve the quality of information contained in extensive databases. This work found that an adequate statistical methodology makes it possible to extract additional information based on existing data.

This study used statistical support as a method of analyzing geochemical data, raising the possibility of using the statistical information obtained to aid in chemostratigraphic correlations. This methodology can be used and expanded in further studies with data from new chemical analysis for the same or another place of study.

Acknowledgments

The authors would like to thank the Postgraduate Program at the Federal University of Rio Grande do Sul for the support of the research. The authors would also like to thank Mr. Álvaro Ramos for the math support and Mr. William Meeks for reviewing the English manuscript and Mrs. Patrícia Zacca for helping with the figures of the study area.

REFERENCES

Aitchison, J. (1982). The statistical analysis of compositional data (with discussion). J.R. Statist. Soc.B44, 139-177.

Aitchison, J. (1986a). The Statistical Analysis of Compositional Data. London: Chapman and Hall. Reprinted in 2003 by The Blackburn Press.

Aitchison, J. (1986b). CODA: A Software Package for Compositional Data (with diskette). London: Chapman and Hall.

Carvalho, P. F. 1932. Geological survey in the State of Rio Grande do Sul. Rio de Janeiro, Geological and Mineralogical Service of Brazil, p. 1-72 (Bulletin 66).

Daunis-I-Estadella, J., Barceló-Vidal, C., Buccianti, A. 2006. Exploratory compositional data analysis. In: Bucianti, A., Mateus-Figueras, G., Pawlowsky-Glahn, V. (eds).
Compositional Data Analysis.

Denalle, H.P. 2013. Geochemistry of the Guaritas Allogroup (Ordovician of the Camaquã Basin, RS. Master's Dissertation, PPGGEO/UFRGS, Porto Alegre, RS. 92p.

Egozcue, J., Pawlowsky-Glahn, V., Mateu-Figueras, G. And Barcélovidal, C., 2003. Isometric log ratio transformations for compositional data analysis. Math. Geol.
35, 41 279-300.

Fleck, Marcelo P.A. E Bourdel, Marie C. Simulation method and choice of factors in principal component analysis. Journal of Public Health [online]. 1998, v. 32, no. 3 [Accessed 11 October 2021], pp. 267-272. Available at: <https://doi.org/10.1590/S0034-89101998000300010>. Epub 07 Aug 2001. ISSN 1518-8787. https://doi.org/10.1590/S0034-89101998000300010.

Fulvio Franchi, Tebogo Kelepile, Andrea Di Capua, Mike CJ De Wit, Onalethata Kemiso, Reneilwe Lasarwe, Octavian Catuneanu. Lithostratigraphy, sedimentary petrography and geochemistry of the Upper Karoo Supergroup in the Central Kalahari Karoo Sub-Basin, Botswana. Journal of African Earth Sciences, Elsevier. January 2021.

Gómez, J.J., Lillo, J. & Sahún, b. Naturally occurring arsenic in groundwater and identification of the geochemical sources in the Duero Cenozoic Basin, Spain. Environ Geol 50, 1151–1170 (2006). https://doi.org/10.1007/s00254-006-0288-z

Hair, J.F.Jr., ANDERSON, R.E., TATHAM, R.L. & BLACK, W.C. Multivariate Data Analysis with Readings. 4th Ed., Prentice Hall, Inc., New Jersey, 1995.

Hartmann, L.A., 1998. Deepest exposed Crust of Brazil – Geochemistry of Paleoproterozoic Depleted Santa Maria Chico Granulites. Gondwana Research, 1:331-441.

Hongyu, K. Comparison of weighted and AMMI-weighted GGEbiplot with other models of genotype × environment interaction. 2015. 155p. Thesis (Doctorate in Statistics and Agronomic Experimentation) - School of Agriculture "Luiz de Queiroz", University of São Paulo, Piracicaba, 2015.

José Leonardo Silva Andriotti. Fundamentals of Statistics and Geostatistics Publisher: UNISINOS (Brazil), 9.165. 2004.

Manly, B.F.J. Multivariate statistical methods. New York, Chapman and Hall, 159 p. 1986.

Maraschin, A.J. et al. (2010) Illite authigenesis in sandstones of the Guaritas Allogroup (Early): Implications for the depositional age, stratigraphy and evolution of the Camaquã Basin (Southern Brazil). Journal of South American Earth Sciences 29, 400-411

Matheron G. Traité de Géostatistique Appliquée, Tome I: Mémoires du Bureau de Recherches Géologiques et Minières. Pairs: Editions Technip, vol. 14, 1962.

Matheron G. Traité de Géostatistique Appliquée, Tome II: Mémoires du Bureau de Recherches Géologiques et Minières. Paris: Editions Technip, vol. 24, 1963.

Matheron G. Les variables regionalisées et leur estimation: an application de la theory des fonctions aléatoires aux sciences de la nature. Paris: Masson et Cie, 1965.

Melo Júnior, J.C.F. de; Sedyama, G.C.; Ferreira, P.A.; Leal, B. G. Determination of homogeneous regions regarding rainfall frequency distribution in eastern Minas Gerais State. Brazilian Journal of Agricultural and Environmental Engineering, v.10, p.408-416, 2006.

PAIM, P.S.G.; CHEMALE JR. F.; LOPES, R. C. 2000. The Camaquã Basin. In: Holz M.; De Ros L. F. (eds.) Geology of Rio Grande do Sul. Porto Alegre, Federal University of Rio Grande do Sul, p. 231-274

Paim, P.S.G. & Scherer, C.M.S. (2002); Stratigraphic architecture of fluvio-wind successions: the example of the Alogrupo Guaritas in the Pedra Pintada region, Rio Grande do Sul, Brazil. In: PSG

Paim; U.F. Faccini; RG Neo (Eds) Geometry, architecture and heterogeneities of sedimentary bodies - Case Studies, São Leopoldo, p. 38-58.

Pawlowsky-Glahn, V., Egozcue, J.J. 2006. Compositional data and their analysis: an introduction. In: Bucianti, A., Mateus-Figueras, G., Pawlowsky-Glahn, V. (eds). Compositional Data Analysis in Geosciences: From Theory to Practice. Geological Society, London, Special Publication 264, 1-10. s

Pinto, L., Alarcón, P., Morton, A., Naipauer, M. "Geoquímica de minerais pesados e geocronologia de zircão detrital U – Pb na Bacia de Manantiales: Implicações para o levantamento da Cordilheira Frontal e conectividade da bacia do foreland nos Andes do centro Argentina". Paleogeografia, Paleoclimatologia, Paleoecologia 492 (2018): 104-125.

Prundeanu, I.M, et al. Geoquímica elementar do reservatório do Cretáceo Superior e formações circunvizinhas aplicada na geo-orientação de poços horizontais, Campo de Lebăda - Mar Negro Ocidental. Oil & Gas Science and Technology - Rev. IFP Energies nouvelles 76, 1 (2021).

R Equipe Central (2020). R: Uma linguagem e ambiente para computação estatística. (Versão 4.0) [Software de computador]. Obtido em https://cran.r-project.org. (Pacotes R recuperados do instantâneo MRAN 2020-08-24). [2]

Ramkumar, Mu. Editor. Conceitos, técnicas e aplicações de quimotratigrafia. Departamento de Geologia, Universidade Periyar, Salem, Tamilnadu, Índia; Laboratório de Pesquisa de Carbonato do Sudeste Asiático (Seacarl), Universiti Teknologi Petronas, Tronoh, Malásia, 2015. ISBN: 978-0-12-419968-2.

Revelle, W. (2019). psych: Procedures for Psychological, Psychometric, and Personality Research. [Pacote R]. Obtido em https://cran.r-project.org/package=psych.[3]

O projeto Jamovi (2021). Jamovi. (Versão 1.6) [Software de computador]. Obtido em https://www.jamovi.org. [1]

Appendix

Statistically significant parameters for the correlation lines shown in the graphs.

1 - Elements: Al2O3 and MnO (Invalid correlation)
Regression line: y = 0.004x-0.01
Beta hypothesis = 0: Not rejected with tcal = 1.83 (95% significance)
Confidence interval for beta (95%): (0.0.01)
Sample linear correlation coefficient: r = 0.35 (positive correlation)
Hypothesis r = 0: Not rejected with tcal = 1.71 (95% significance)

2 - Elements: Al2O3 and Fe2O3T (Valid Correlation)
Regression line: y = 0.26x-1.02
Beta Hypothesis = 0: Rejected with tcal = 2.47 (95% Significance)
Confidence interval for beta (95%): (0.0.53)
Sample linear correlation coefficient: r = 0.45 (positive correlation)
Hypothesis r = 0: Not rejected with tcal = 2.2 (95% significance)

3 - Elements: Fe2O3T and CaO (Invalid correlation)
Regression line: y = 0.66x+0.13
Beta hypothesis = 0: Not rejected with tcal = 1.71 (95% significance)
Confidence interval for beta (95%): (-0.3,1.61)
Sample linear correlation coefficient: r = 0.33 (positive correlation)
Hypothesis r = 0: Not rejected with tcal = 1.62 (95% significance)

4 - Elements: Fe2O3T and TiO2 (Valid Correlation)
Regression line: y = 0.06x+0.04
Beta Hypothesis = 0: Rejected with tcal = 3.78 (95% Significance)
Confidence interval for beta (95%): (0.02.0.1)
Sample linear correlation coefficient: r = 0.61 (positive correlation)
Hypothesis r = 0: Rejected with tcal = 2.99 (95% significance)

5 - Elements: SiO2 and Na2O (Valid Correlation)
Regression line: y = -0.06x+6.42
Beta Hypothesis = 0: Rejected with tcal = -6.54 (95% Significance)
Confidence interval for beta (95%): (-0.08,-0.04)
Sample linear correlation coefficient: r = -0.8 (negative correlation)
Hypothesis r = 0: Rejected with tcal = -3.92 (95% significance)

6 - Elements: SiO2 and CaO (Valid Correlation)
Regression line: y = -0.27x+22.48
Beta Hypothesis = 0: Rejected with tcal = -5.3 (95% Significance)
Confidence interval for beta (95%): (-0.39,-0.14)
Sample linear correlation coefficient: r = -0.73 (negative correlation)
Hypothesis r = 0: Rejected with tcal = -3.6 (95% significance)

7 - Elements: SiO2 and K2O (Invalid Correlation)
Regression line: y = -0.01x+4.92
Beta Hypothesis = 0: Not rejected with tcal = -0.84 (95% Significance)
Confidence interval for beta (95%): (-0.06.0.03)
Sample linear correlation coefficient: r = -0.17 (weak correlation)
Hypothesis r = 0: Not rejected with tcal = -0.83 (95% significance)

8 - Elements: SiO2 and Zr (Valid Correlation)
Regression line: y = -8.81x+799.87
Beta Hypothesis = 0: Rejected with tcal = -3.6 (95% Significance)
Confidence interval for beta (95%): (-14.92,-2.71)
Sample linear correlation coefficient: r = -0.59 (negative correlation)
Hypothesis r = 0: Rejected with tcal = -2.9 (95% significance)

9 - Elements: Sr and Rb (Invalid Correlation)
Regression line: y = 0.05x+134.47
Beta Hypothesis = 0: Not rejected with tcal = 0.53 (95% Significance)
Confidence interval for beta (95%): (-0.19.0.29)
Sample linear correlation coefficient: r = 0.11 (weak correlation)
Hypothesis r = 0: Not rejected with tcal = 0.52 (95% significance)

10 - Elements: Sr and Ba (Invalid Correlation)
Regression line: y = 2.44x+451.63
Beta hypothesis = 0: Not rejected with tcal = 0.76 (95% Significance)
Confidence interval for beta (95%): (-5.52,10.41)
Sample linear correlation coefficient: r = 0.15 (weak correlation)
Hypothesis r = 0: Not rejected with tcal = 0.75 (95% significance)

11 - Elements: Ba and Rb (Invalid Correlation)
Regression line: y = -0.01x+149.74
Beta Hypothesis = 0: Rejected with tcal = -2.82 (95% Significance)
Confidence interval for beta (95%): (-0.03.0)
Sample linear correlation coefficient: r = -0.5 (negative correlation)
Hypothesis r = 0: Not rejected with tcal = -2.44 (95% significance)

12 - Elements: CaO and Sr (Invalid Correlation)
Regression line: y = 6.9x+93.14
Beta hypothesis = 0: Not rejected with tcal = 1.74 (95% significance)

Confidence interval for beta (95%): (-2.98.16.78)
Sample linear correlation coefficient: r = 0.33 (positive correlation)
Hypothesis r = 0: Not rejected with tcal = 1.64 (95% significance)

13 - Elements: CaO and SiO2 (Valid Correlation)
Regression line: y = -2.01x+82.15
Beta Hypothesis = 0: Rejected with tcal = -5.3 (95% Significance)
Confidence interval for beta (95%): (-2.96,-1.07)
Sample linear correlation coefficient: r = -0.73 (negative correlation)
Hypothesis r = 0: Rejected with tcal = -3.6 (95% significance)

14 - Elements: CaO and K2O (Invalid Correlation)
Regression line: y = -0.05x+3.79
Beta Hypothesis = 0: Not rejected with tcal = -0.99 (95% Significance)
Confidence interval for beta (95%): (-0.17.0.07)
Sample linear correlation coefficient: r = -0.2 (weak correlation)
Hypothesis r = 0: Not rejected with tcal = -0.97 (95% significance)

15 - Elements: CaO and MgO (Invalid Correlation)
Regression line: y = 0.0026x+0.33
Beta Hypothesis = 0: Not rejected with tcal = 0.14 (95% Significance)
Confidence interval for beta (95%): (-0.04.0.05)
Sample linear correlation coefficient: r = 0.03 (weak correlation)
Hypothesis r = 0: Not rejected with tcal = 0.14 (95% significance)

16 - Elements: SiO2 and MgO (Valid Correlation)
Regression line: y = -0.02x+1.95
Beta Hypothesis = 0: Rejected with tcal = -3.73 (95% Significance)
Confidence interval for beta (95%): (-0.03,-0.01)
Sample linear correlation coefficient: r = -0.61 (negative correlation)
Hypothesis r = 0: Rejected with tcal = -2.97 (95% significance)

17 - Elements: CaO and Na2O (Valid Correlation)
Regression line: y = 0.12x+1.44
Beta Hypothesis = 0: Rejected with tcal = 3.39 (95% Significance)
Confidence interval for beta (95%): (0.03.0.21)
Sample linear correlation coefficient: r = 0.57 (positive correlation)
Hypothesis r = 0: Rejected with tcal = 2.79 (95% significance)

18 - Elements: SiO2 and Al2O3 (Valid Correlation)
Regression line: y = -0.25x+29.02
Beta hypothesis = 0: Rejected with tcal = -6.24 (Significance of 5%)
Confidence interval for beta (95%): (-0.35,-0.15)
Sample linear correlation coefficient: r = -0.79 (negative correlation)
Hypothesis r = 0: Rejected with tcal = -3.85 (5% significance)

7. Referências

Aitchison, J. (1982). **The statistical analysis of compositional data (with discussion).** J. R. Statist. Soc.B44, 139-177.

Aitchison, J. (1986a). **The Statistical Analysis of Compositional Data. London: Chapman and Hall.** Reprinted in 2003 by The Blackburn Press.

Aitchison, J. (1986b). **CODA: A Software Package for Compositional Data (with diskette).** London: Chapman and Hall.

Carvalho, P. F. 1932. **Reconhecimento geológico no Estado do Rio Grande do Sul. Rio de Janeiro, Serviço Geológico e Mineralógico do Brasil,** p. 1-72 (Boletim 66).

Daunis-I-Estadella, J., Barceló-Vidal, C., Buccianti, A. 2006. **Exploratory compositional data analysis.** In: Bucianti, A., Mateus-Figueras, G., Pawlowsky-Glahn, V. (eds). Compositional Data Analysis.

Denalle, H.P. 2013. **Geoquímica do Alogrupo Guaritas (Ordoviciano da Bacia do Camaquã, RS.** Dissertação de Mestrado, PPGGEO/UFRGS, Porto Alegre, RS. 92p.

Egozcue, J., Pawlowsky-Glahn, V., Mateu-Figueras, G. And Barcélovidal, C., 2003. **Isometric log ratio transformations for compositional data analysis. Math.** Geol. 35, 41 279-300.

Fleck, Marcelo P.A. E Bourdel, Marie C. **Método de simulação e escolha de fatores na análise dos principais componentes. Revista de Saúde Pública** [online]. 1998, v. 32, n. 3 [Acessado 11 Outubro 2021] , pp. 267-272. Disponível em: <https://doi.org/10.1590/S0034-89101998000300010>. Epub 07 Ago 2001. ISSN 1518-8787. https://doi.org/10.1590/S0034-89101998000300010.

Fulvio Franchi, Tebogo Kelepile, Andrea Di Capua,Mike C.J. De Wit, Onalethata Kemiso, Reneilwe Lasarwe, Octavian Catuneanu. **Lithostratigraphy, sedimentary petrography and geochemistry of the Upper Karoo Supergroup in the Central Kalahari Karoo Sub-Basin, Botswana.** Journal of African Earth Sciences, Elsevier. January 2021.

GeoGebra, http://www.geogebra.org[1]

Gómez, J.J., Lillo, J. & Sahún, b. **Naturally occurring arsenic in groundwater and identification of the geochemical sources in the Duero Cenozoic Basin, Spain**. Environ Geol 50, 1151–1170 (2006). https://doi.org/10.1007/s00254-006-0288-z

Hair, J. F. Jr., ANDERSON, R. E., TATHAM, R. L. & BLACK, W. C. **Multivariate Data Analysis with Readings**. 4ª Ed., Prentice Hall, Inc., New Jersey, 1995.

Hartmann, L.A., 1998. **Deepest exposed Crust of Brazil – Geochemestry of Paleoproterozoic Depleted Santa Maria Chico Granulites.** Gondwana Research, 1: 331 -441.

Hongyu, K. **Comparação do GGEbiplot ponderado e AMMI-ponderado com outros modelos de interação genótipo × ambiente.** 2015. 155p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2015.

José Leonardo Silva Andriotti. **Fundamentos De Estatística E Geoestátistica** Editora: UNISINOS (Brasil), 9.165. 2003.

Manly, B. F. J. **Multivariate statistical methods**. New York, Chapman and Hall, 159 p. 1986.

Maraschin, A.J. et al. (2010) **Illite authigenesis in sandstones of the Guaritas Allogroup (Early): Implications for the depositional age, stratigraphy and evolution of the Camaquã Basin (Southern Brazil).** Journal of South American Earth Sciences 29, 400-411

Matheron G. **Traité de Géostatistique Appliquée, Tome I: Mémoires du Bureau de Recherches Géologiques et Minières.** Pairs: Editions Technip, vol. 14, 1962.

Matheron G. **Traité de Géostatistique Appliquée, Tome II: Mémoires du Bureau de Recherches Géologiques et Minières.** Paris: Editions Technip, vol. 24, 1963.

Matheron G. **Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature.** Paris: Masson et Cie, 1965.

Melo Júnior, J. C. F. de; Sedyama, G. C.; Ferreira, P. A.; Leal, B. G. **Determinação de regiões homogêneas quanto à distribuição de frequência de chuvas no leste do Estado de Minas Gerais.** Revista Brasileira de Engenharia Agrícola e Ambiental, v.10, p.408-416, 2006.

Paim, P.S.G.; CHEMALE JR. F.; LOPES, R. C. 2000. **A Bacia do Camaquã. In: Holz M.; De Ros L. F. (eds.) Geologia do Rio Grande do Sul**. Porto Alegre, Universidade Federal do rio Grande do Sul, p. 231-274

Paim, P. S. G. & Scherer, C. M. S. (2002); **Arquitetura estratigráfica de sucessões flúvio-eólicas: o exemplo do Alogrupo Guaritas na região da Pedra Pintada, Rio Grande do Sul, Brasil.** In: P. S. G. Paim; U. F. Faccini; R. G. Neo (Eds) Geometria, arquitetura e heterogeneidades de corpos sedimentares - Estudo de Casos, São Leopoldo, p. 38-58.

Pawlowsky-Glahn, V., Egozcue, J. J. 2006. **Compositional data and their analysis: an introduction. In: Bucianti, A., Mateus-Figueras, G., Pawlowsky-Glahn, V. (eds). Compositional Data Analysis in Geosciences: From Theory to Practice.** Geological Society, London, Special Publication 264, 1-10. S

Pinto, L., Alarcón, P., Morton, A., Naipauer, M. "**Geochemistry of heavy minerals and U–Pb detrital zircon geochronology in the Manantiales Basin: Implications for Frontal Cordillera uplift and foreland basin connectivity in the Andes of central Argentina**". *Palaeogeography, Palaeoclimatology, Palaeoecology* 492 (2018) : 104-125.

Prundeanu, I.M, et al. **Elemental geochemistry of the Upper Cretaceous reservoir and surrounding formations applied in geosteering of horizontal wells, Lebăda Field – Western Black Sea. Oil & Gas Science and Technology** – Rev. IFP Energies nouvelles 76, 1 (2021).

R Core Team (2020). **R: A Language and environment for statistical computing. (Version 4.0) [Computer software].** Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2020-08-24).

Ramkumar, Mu. Editor. **Chemostratigraphy Concepts, Techniques, and Applications.** Department of Geology, Periyar University, Salem, Tamilnadu, India; South East Asia Carbonate Research Laboratory (Seacarl), Universiti Teknologi Petronas, Tronoh, Malaysia, 2015. ISBN: 978-0-12-419968-2.

The jamovi project (2021). jamovi. (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org.[2]

8. Considerações Finais

O estudo propôs uma metodologia que pode ser desenvolvida e aplicável a ambientes análogos com similaridades quanto à composição química das rochas e os processos envolvidos em sua deposição como as Aloformações Pedra Pintada e Varzinha do Alogrupo Guaritas da Bacia do Camaquã (RS).

Assim, o emprego de métodos estatísticos mostrou-se vantajoso em estudos de estratigrafia química, sobretudo, quando já existem dados disponibilizados. Mesmo para uma área em que ainda inexistem análises, o tratamento estatístico para os dados de novas análises confere benefícios ao passo que pode restringir o número de elementos alvo do estudo, reduzindo a necessidade de sucessivas análises químicas adicionais mais abrangentes, que envolvam mais de 50 elementos químicos distintos. Evidentemente que isso tudo a depender dos objetivos da pesquisa e interesse por certos minerais e composições químicas específicas.

Desse modo, para futuras pesquisas de correlação quimioestratigráfica nesta área de estudo poderá ser possível realizar análises químicas de rocha total para um menor número de elementos menores e traço. Portanto, com a PCA foi possível estabelecer com significância estatística quais composições químicas são mais distintivas para as regiões amostradas ao longo do perfil colunar no Alogrupo Guaritas.

Além da vantagem de uma possível redução de custos com análises químicas, os métodos estatísticos são capazes de aprimorar a qualidade da informação contida em extensos bancos de dados. Esse trabalho verificou que uma metodologia estatística adequada possibilita extrair informação adicional com base nos dados existentes.

Este estudo utilizou assim o suporte estatístico como método de análise dos dados geoquímicos, levantando a possibilidade de utilização da informação estatística obtida para auxílio em correlações quimioestratigráficas. Metodologia essa que pode ser utilizada e ampliada em estudos posteriores com dados de novas análises químicas para o mesmo ou outro local de estudo.

Apêndice

Parâmetros calculados de significância estatística para as retas de correlação apresentadas nos gráficos.

1 - Elementos: Al_2O_3 e MnO (Correlação inválida)
Reta de regressão: y = 0,004x-0,01
Hipótese beta = 0: Não rejeitada com tcal = 1,83 (Significância de 95%)
Intervalo de confiança para beta (95%): (0,0,01)
Coeficiente de correlação linear amostral: r = 0,35 (correlação positiva)
Hipótese r = 0: Não rejeitada com tcal = 1,71 (significância de 95%)

2 - Elementos: Al_2O_3 e Fe_2O_3T (Correlação válida)
Reta de regressão: y = 0,26x-1,02
Hipótese beta = 0: Rejeitada com tcal = 2,47 (Significância de 95%)
Intervalo de confiança para beta (95%): (0,0,53)
Coeficiente de correlação linear amostral: r = 0,45 (correlação positiva)
Hipótese r = 0: Não rejeitada com tcal = 2,2 (significância de 95%)

3 - Elementos: Fe_2O_3T e CaO (Correlação inválida)
Reta de regressão: y = 0,66x+0,13
Hipótese beta = 0: Não rejeitada com tcal = 1,71 (Significância de 95%)
Intervalo de confiança para beta (95%): (-0,3,1,61)
Coeficiente de correlação linear amostral: r = 0,33 (correlação positiva)
Hipótese r = 0: Não rejeitada com tcal = 1,62 (significância de 95%)

4 - Elementos: Fe_2O_3T e TiO_2 (Correlação válida)
Reta de regressão: y = 0,06x+0,04
Hipótese beta = 0: Rejeitada com tcal = 3,78 (Significância de 95%)
Intervalo de confiança para beta (95%): (0,02,0,1)
Coeficiente de correlação linear amostral: r = 0,61 (correlação positiva)
Hipótese r = 0: Rejeitada com tcal = 2,99 (significância de 95%)

5 - Elementos: SiO_2 e Na_2O (Correlação válida)
Reta de regressão: y = -0,06x+6,42
Hipótese beta = 0: Rejeitada com tcal = -6,54 (Significância de 95%)
Intervalo de confiança para beta (95%): (-0,08,-0,04)

Coeficiente de correlação linear amostral: r = -0,8 (correlação negativa)
Hipótese r = 0: Rejeitada com tcal = -3,92 (significância de 95%)

6 - Elementos: SiO2 e CaO (Correlação válida)
Reta de regressão: y = -0,27x+22,48
Hipótese beta = 0: Rejeitada com tcal = -5,3 (Significância de 95%)
Intervalo de confiança para beta (95%): (-0,39,-0,14)
Coeficiente de correlação linear amostral: r = -0,73 (correlação negativa)
Hipótese r = 0: Rejeitada com tcal = -3,6 (significância de 95%)

7 - Elementos: SiO2 e K2O (Correlação inválida)
Reta de regressão: y = -0,01x+4,92
Hipótese beta = 0: Não rejeitada com tcal = -0,84 (Significância de 95%)
Intervalo de confiança para beta (95%): (-0,06,0,03)
Coeficiente de correlação linear amostral: r = -0,17 (correlação fraca)
Hipótese r = 0: Não rejeitada com tcal = -0,83 (significância de 95%)

8 - Elementos: SiO2 e Zr (Correlação válida)
Reta de regressão: y = -8,81x+799,87
Hipótese beta = 0: Rejeitada com tcal = -3,6 (Significância de 95%)
Intervalo de confiança para beta (95%): (-14,92,-2,71)
Coeficiente de correlação linear amostral: r = -0,59 (correlação negativa)
Hipótese r = 0: Rejeitada com tcal = -2,9 (significância de 95%)

9 - Elementos: Sr e Rb (Correlação inválida)
Reta de regressão: y = 0,05x+134,47
Hipótese beta = 0: Não rejeitada com tcal = 0,53 (Significância de 95%)
Intervalo de confiança para beta (95%): (-0,19,0,29)
Coeficiente de correlação linear amostral: r = 0,11 (correlação fraca)
Hipótese r = 0: Não rejeitada com tcal = 0,52 (significância de 95%)

10 - Elementos: Sr e Ba (Correlação inválida)
Reta de regressão: y = 2,44x+451,63
Hipótese beta = 0: Não rejeitada com tcal = 0,76 (Significância de 95%)
Intervalo de confiança para beta (95%): (-5,52,10,41)
Coeficiente de correlação linear amostral: r = 0,15 (correlação fraca)
Hipótese r = 0: Não rejeitada com tcal = 0,75 (significância de 95%)

11 - Elementos: Ba e Rb (Correlação inválida)
Reta de regressão: y = -0,01x+149,74
Hipótese beta = 0: Rejeitada com tcal = -2,82 (Significância de 95%)

Intervalo de confiança para beta (95%): (-0,03,0)
Coeficiente de correlação linear amostral: r = -0,5 (correlação negativa)
Hipótese r = 0: Não rejeitada com tcal = -2,44 (significância de 95%)

12 - Elementos: CaO e Sr (Correlação inválida)
Reta de regressão: y = 6,9x+93,14
Hipótese beta = 0: Não rejeitada com tcal = 1,74 (Significância de 95%)
Intervalo de confiança para beta (95%): (-2,98,16,78)
Coeficiente de correlação linear amostral: r = 0,33 (correlação positiva)
Hipótese r = 0: Não rejeitada com tcal = 1,64 (significância de 95%)

13 - Elementos: CaO e SiO2 (Correlação válida)
Reta de regressão: y = -2,01x+82,15
Hipótese beta = 0: Rejeitada com tcal = -5,3 (Significância de 95%)
Intervalo de confiança para beta (95%): (-2,96,-1,07)
Coeficiente de correlação linear amostral: r = -0,73 (correlação negativa)
Hipótese r = 0: Rejeitada com tcal = -3,6 (significância de 95%)

14 - Elementos: CaO e K2O (Correlação inválida)
Reta de regressão: y = -0,05x+3,79
Hipótese beta = 0: Não rejeitada com tcal = -0,99 (Significância de 95%)
Intervalo de confiança para beta (95%): (-0,17,0,07)
Coeficiente de correlação linear amostral: r = -0,2 (correlação fraca)
Hipótese r = 0: Não rejeitada com tcal = -0,97 (significância de 95%)

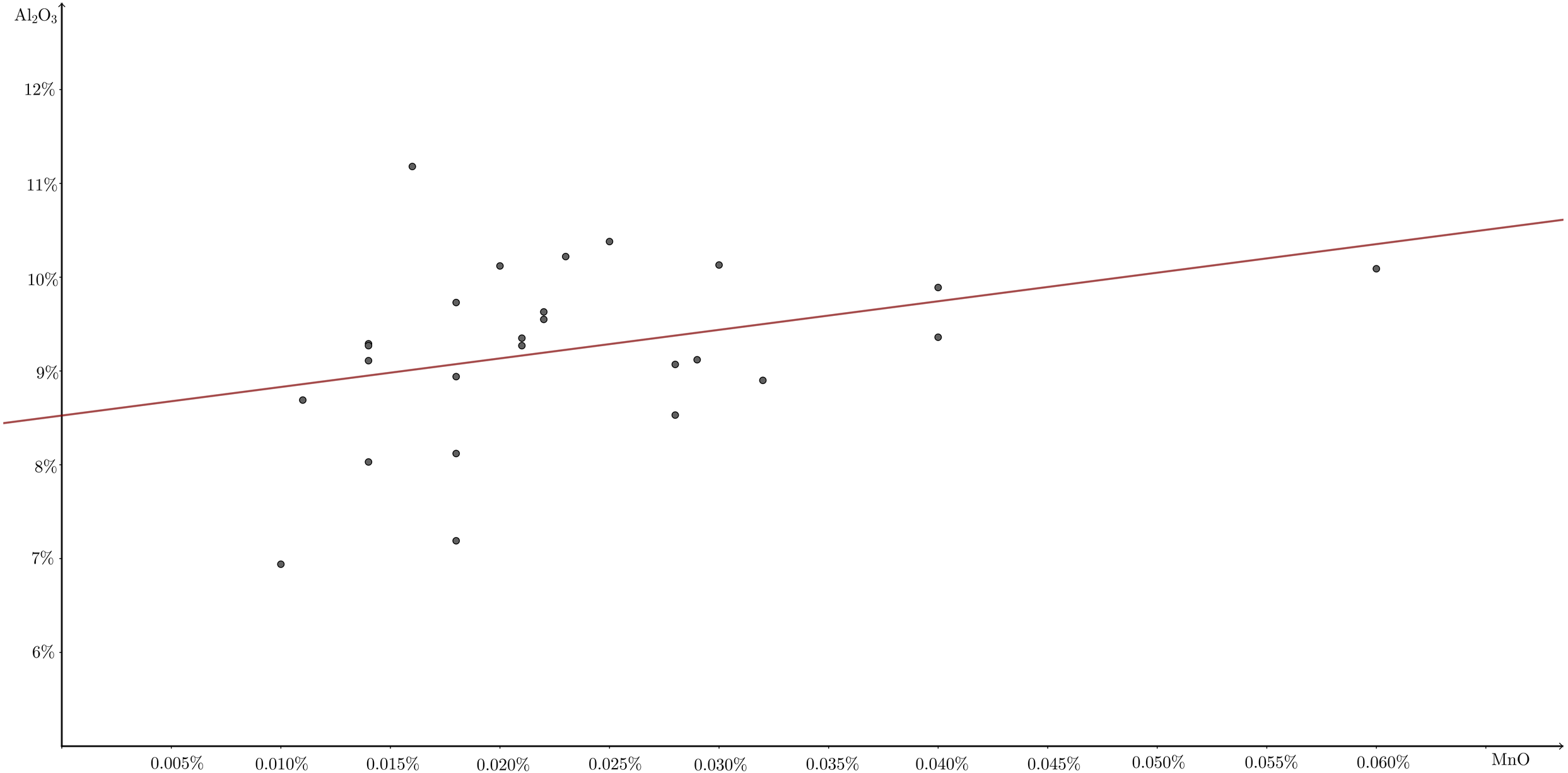
15 - Elementos: CaO e MgO (Correlação inválida)
Reta de regressão: y = 0,0026x+0,33
Hipótese beta = 0: Não rejeitada com tcal = 0,14 (Significância de 95%)
Intervalo de confiança para beta (95%): (-0,04,0,05)
Coeficiente de correlação linear amostral: r = 0,03 (correlação fraca)
Hipótese r = 0: Não rejeitada com tcal = 0,14 (significância de 95%)

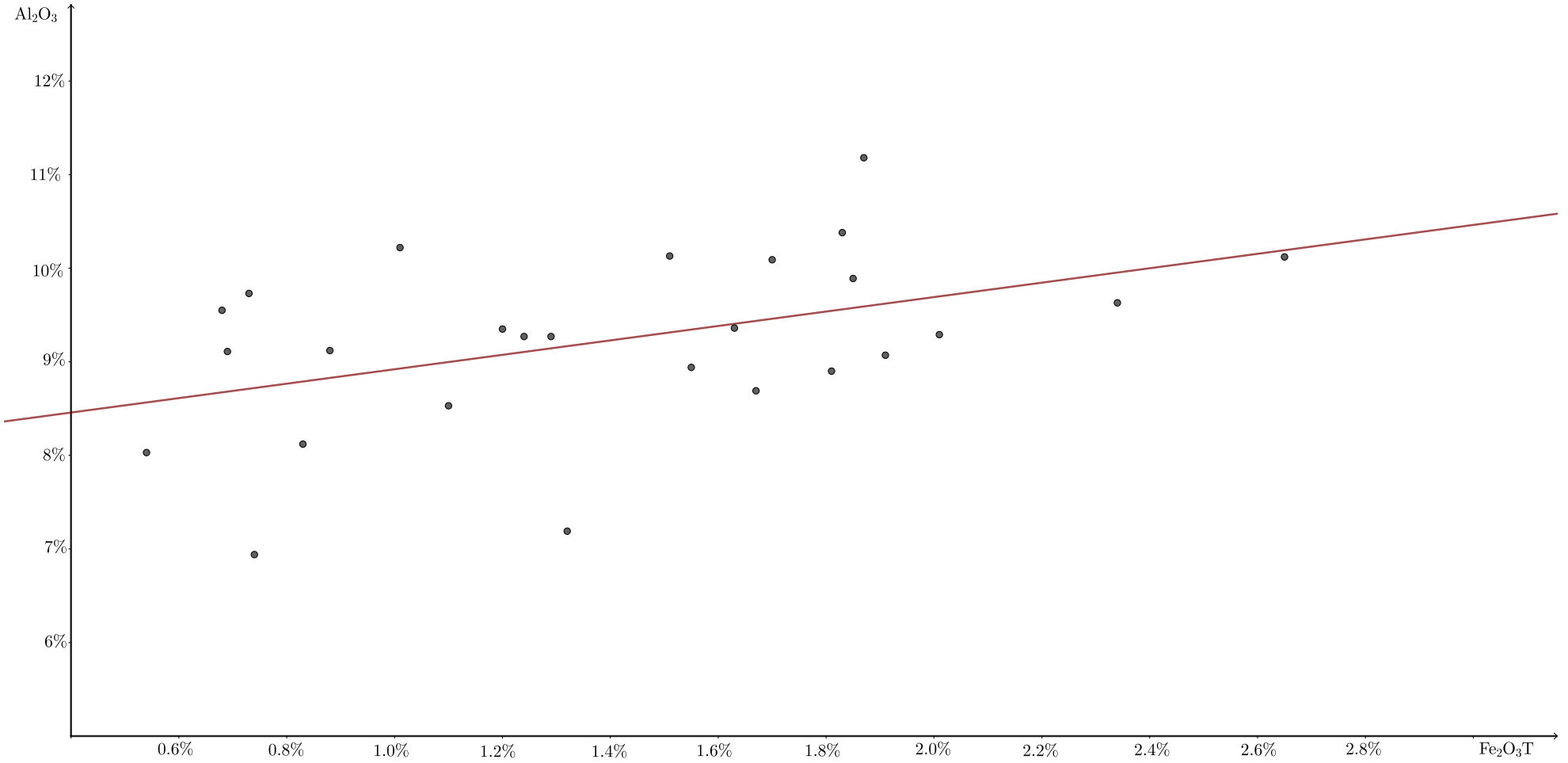
16 - Elementos: SiO2 e MgO (Correlação válida)
Reta de regressão: y = -0,02x+1,95
Hipótese beta = 0: Rejeitada com tcal = -3,73 (Significância de 95%)
Intervalo de confiança para beta (95%): (-0,03,-0,01)
Coeficiente de correlação linear amostral: r = -0,61 (correlação negativa)
Hipótese r = 0: Rejeitada com tcal = -2,97 (significância de 95%)

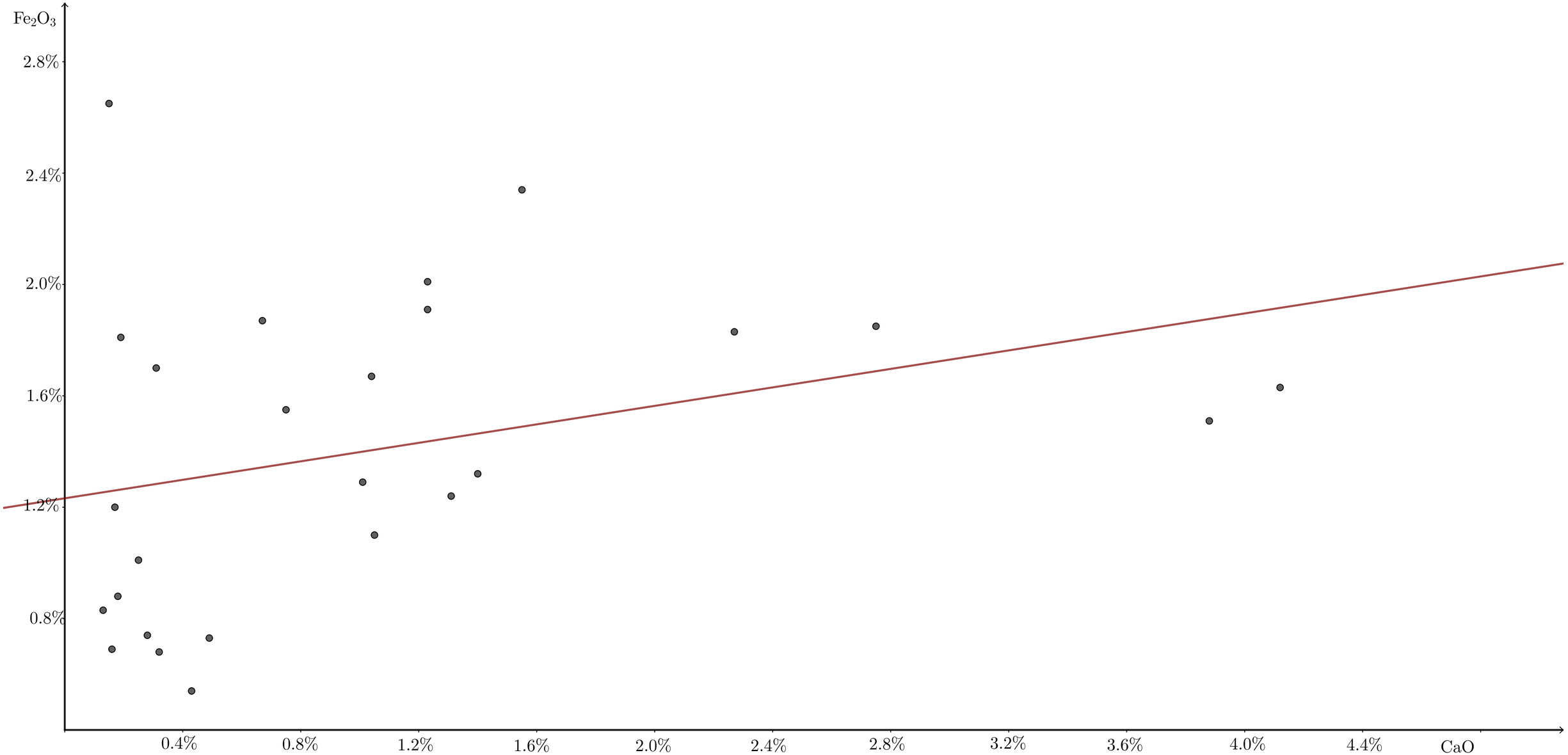
17 - Elementos: CaO e Na2O (Correlação válida)
Reta de regressão: y = 0,12x+1,44

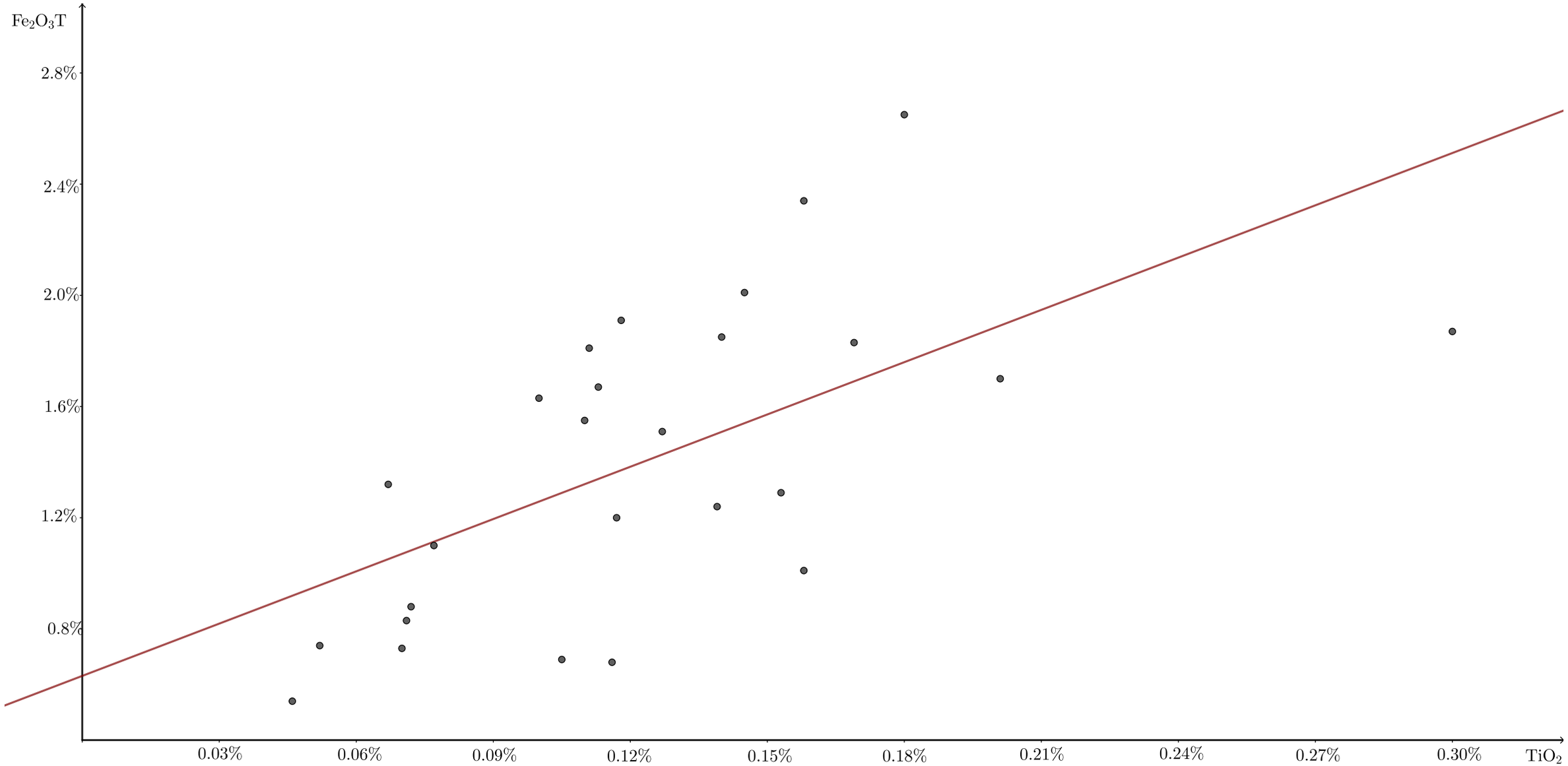
Hipótese beta = 0: Rejeitada com tcal = 3,39 (Significância de 95%)
Intervalo de confiança para beta (95%): (0,03,0,21)
Coeficiente de correlação linear amostral: r = 0,57 (correlação positiva)
Hipótese r = 0: Rejeitada com tcal = 2,79 (significância de 95%)

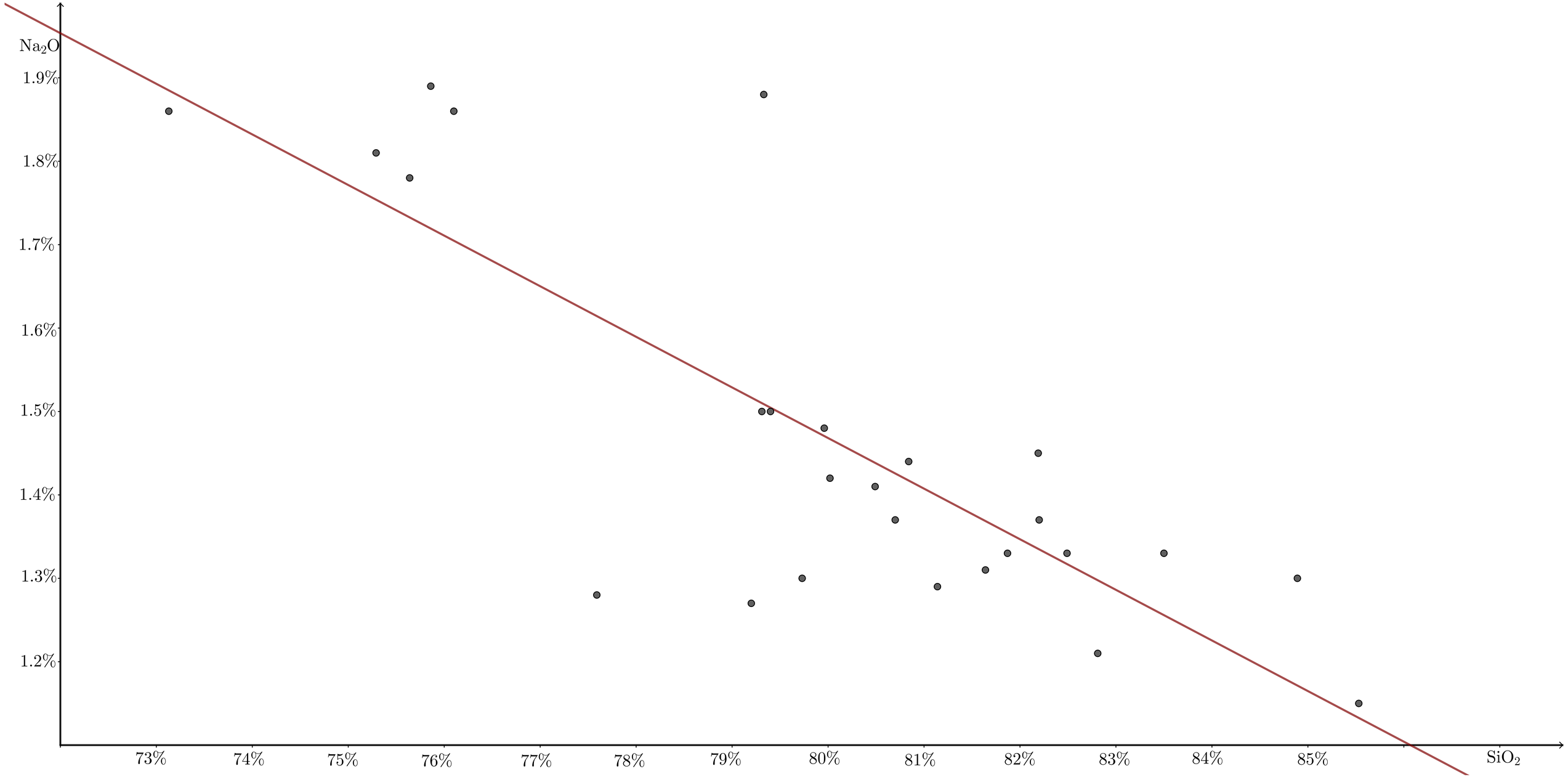
18 - Elementos: SiO_2 e Al_2O_3 (Correlação válida)
Reta de regressão: y = -0,25x+29,02
Hipótese beta = 0: Rejeitada com tcal = -6,24 (Significância de 5%)
Intervalo de confiança para beta (95%): (-0,35,-0,15)
Coeficiente de correlação linear amostral: r = -0,79 (correlação negativa)
Hipótese r = 0: Rejeitada com tcal = -3,85 (significância de 5%)

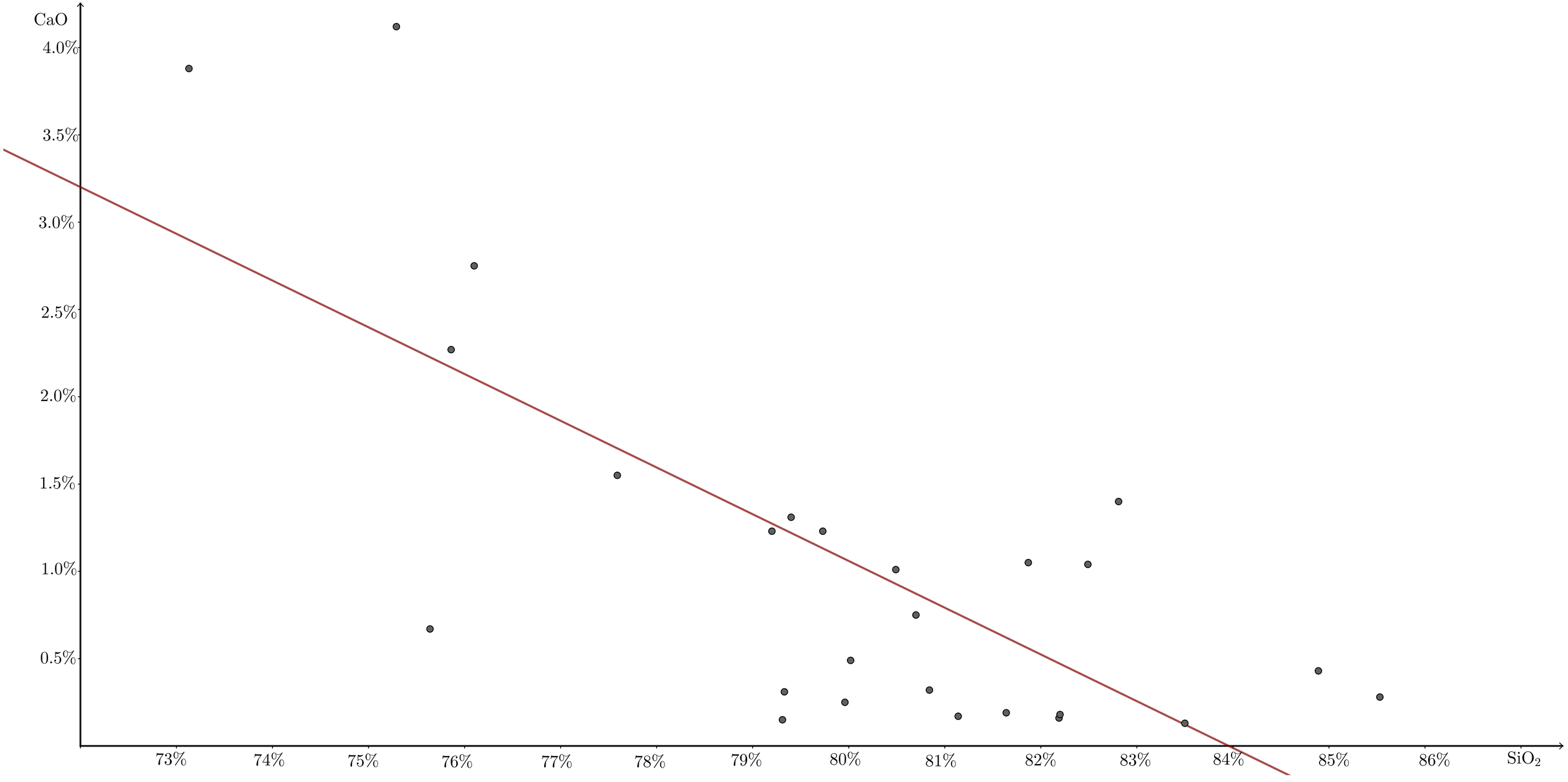

Al_2O_3
12%
11%
10%
9%
8%
7%
6%
0.005%
0.010%
0.015%
0.020%
0.025%
0.030%
0.035%
0.040%
0.045%
0.050%
0.055%
0.060%
MnO

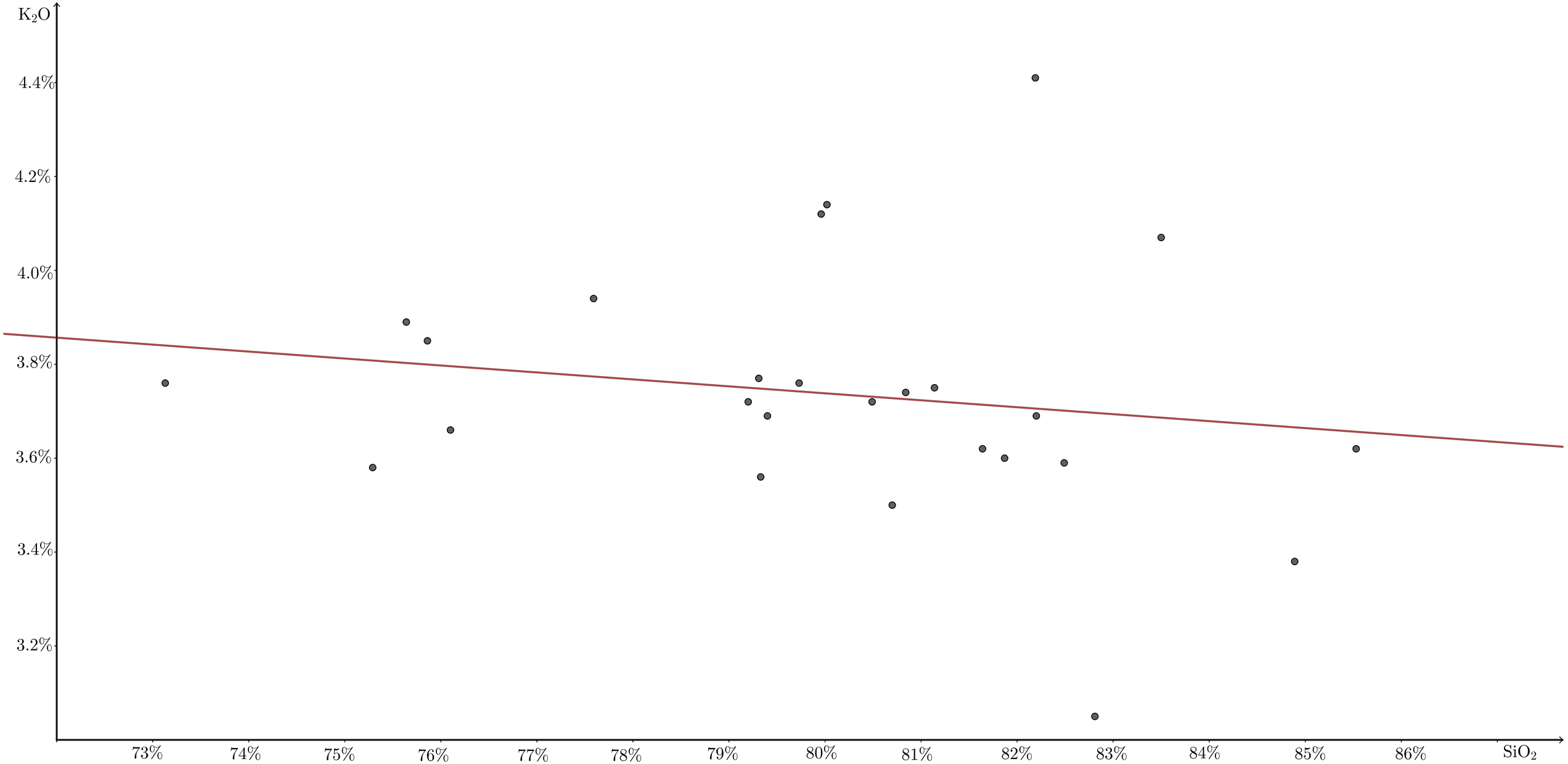

Al_2O_3
12%
11%
10%
9%
8%
7%
6%
0.6%
0.8%
1.0%
1.2%
1.4%
1.6%
1.8%
2.0%
2.2%
2.4%
2.6%
2.8%
Fe_2O_3T

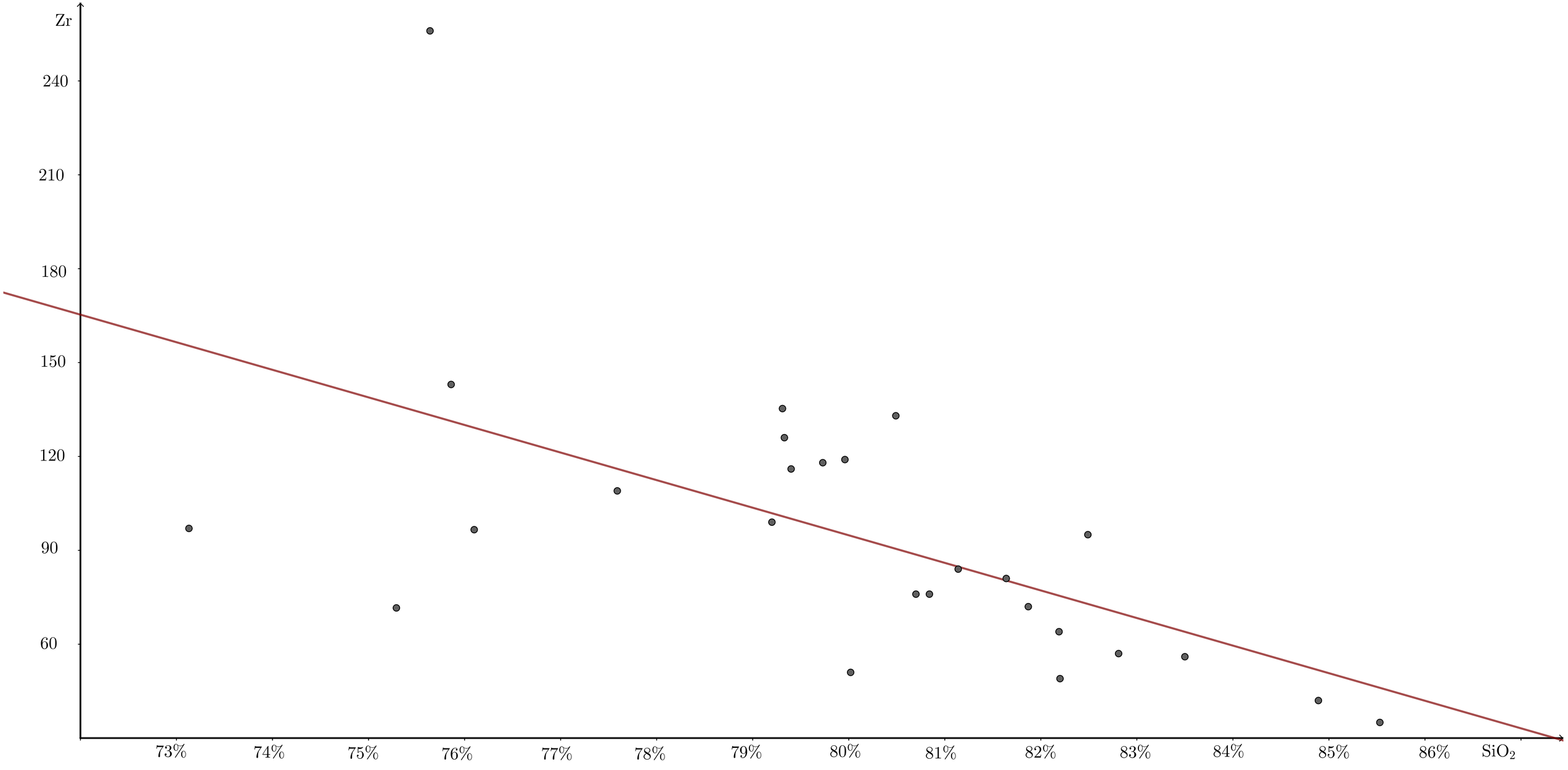

Fe_2O_3
2.8%
2.4%
2.0%
1.6%
1.2%
0.8%
0.4%
0.8%
1.2%
1.6%
2.0%
2.4%
2.8%
3.2%
3.6%
4.0%
4.4%
CaO

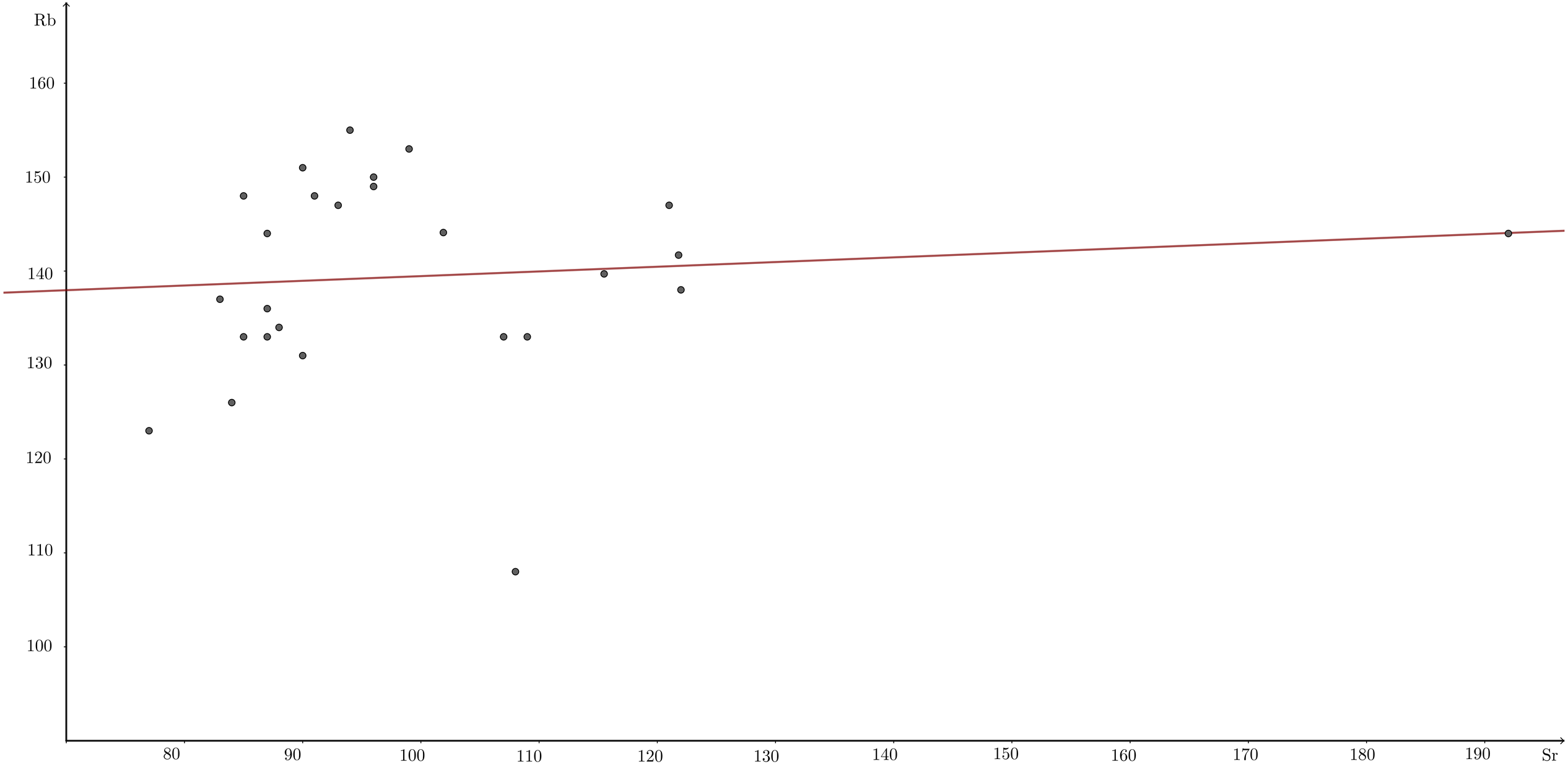

Fe_2O_3T
2.8%
2.4%
2.0%
1.6%
1.2%
0.8%
0.03%
0.06%
0.09%
0.12%
0.15%
0.18%
0.21%
0.24%
0.27%
0.30%
TiO_2

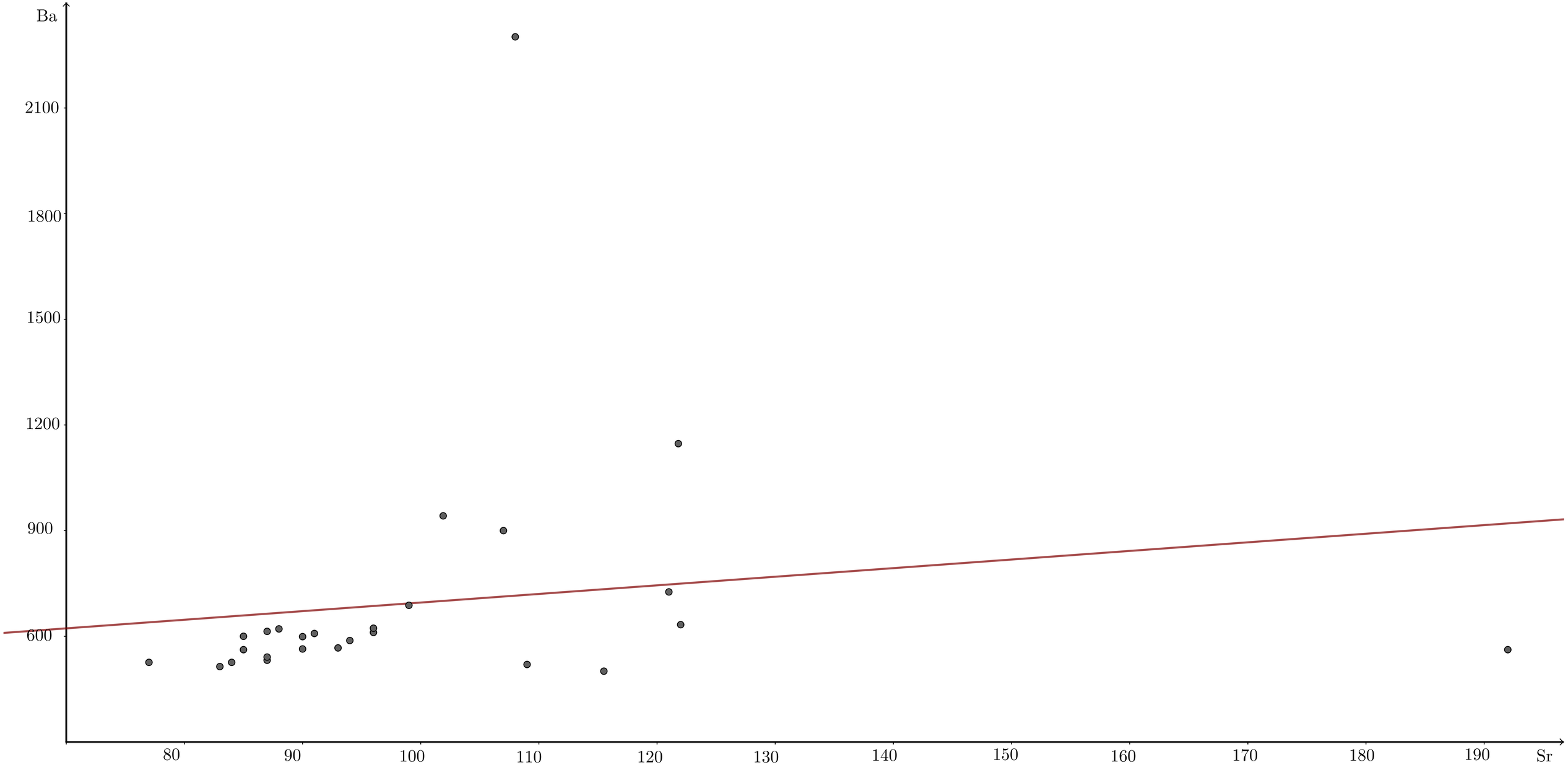

Na_2O
1.9%
1.8%
1.7%
1.6%
1.5%
1.4%
1.3%
1.2%
73%
74%
75%
76%
77%
78%
79%
80%
81%
82%
83%
84%
85%
SiO_2

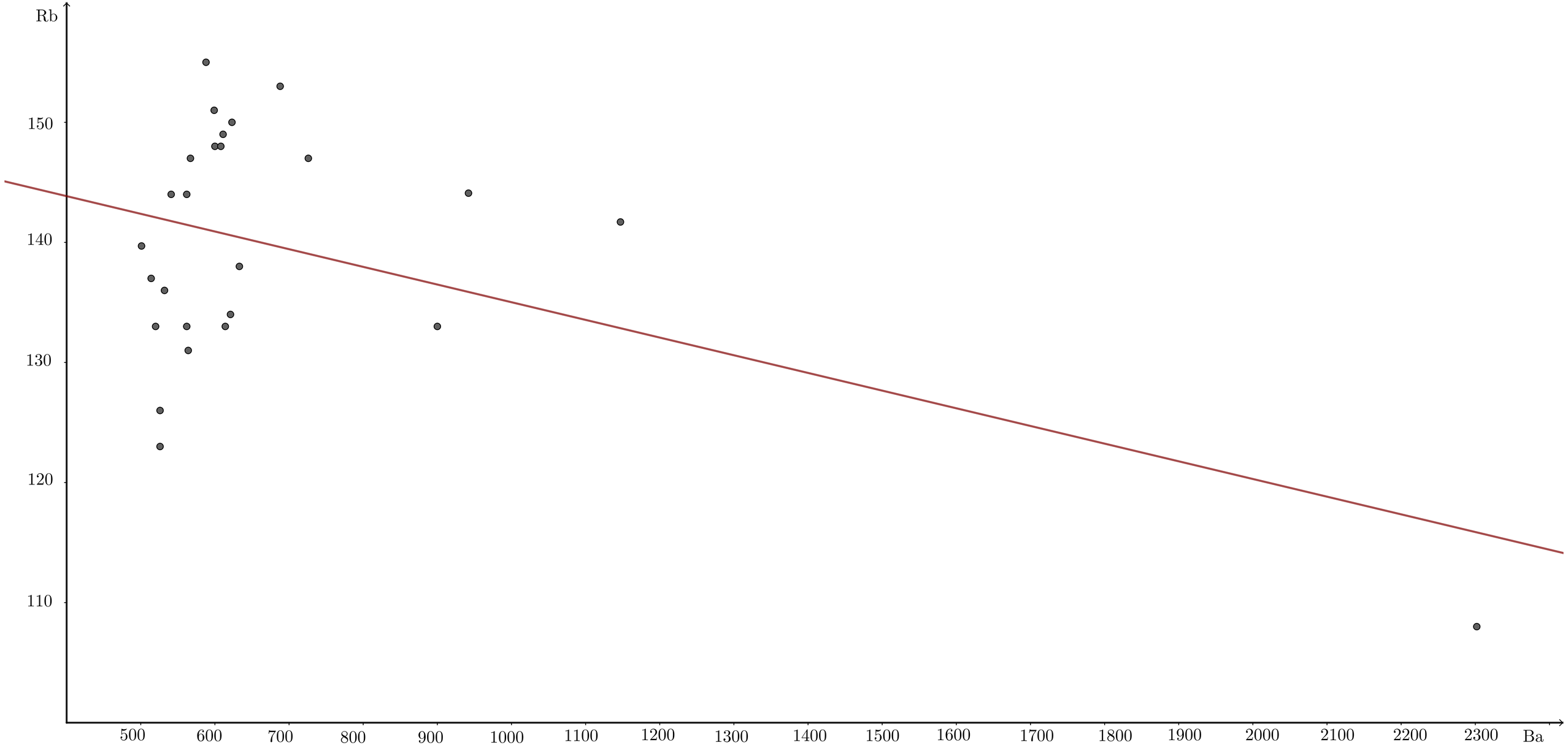

CaO
4.0%
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
73%
74%
75%
76%
77%
78%
79%
80%
81%
82%
83%
84%
85%
86%
SiO2

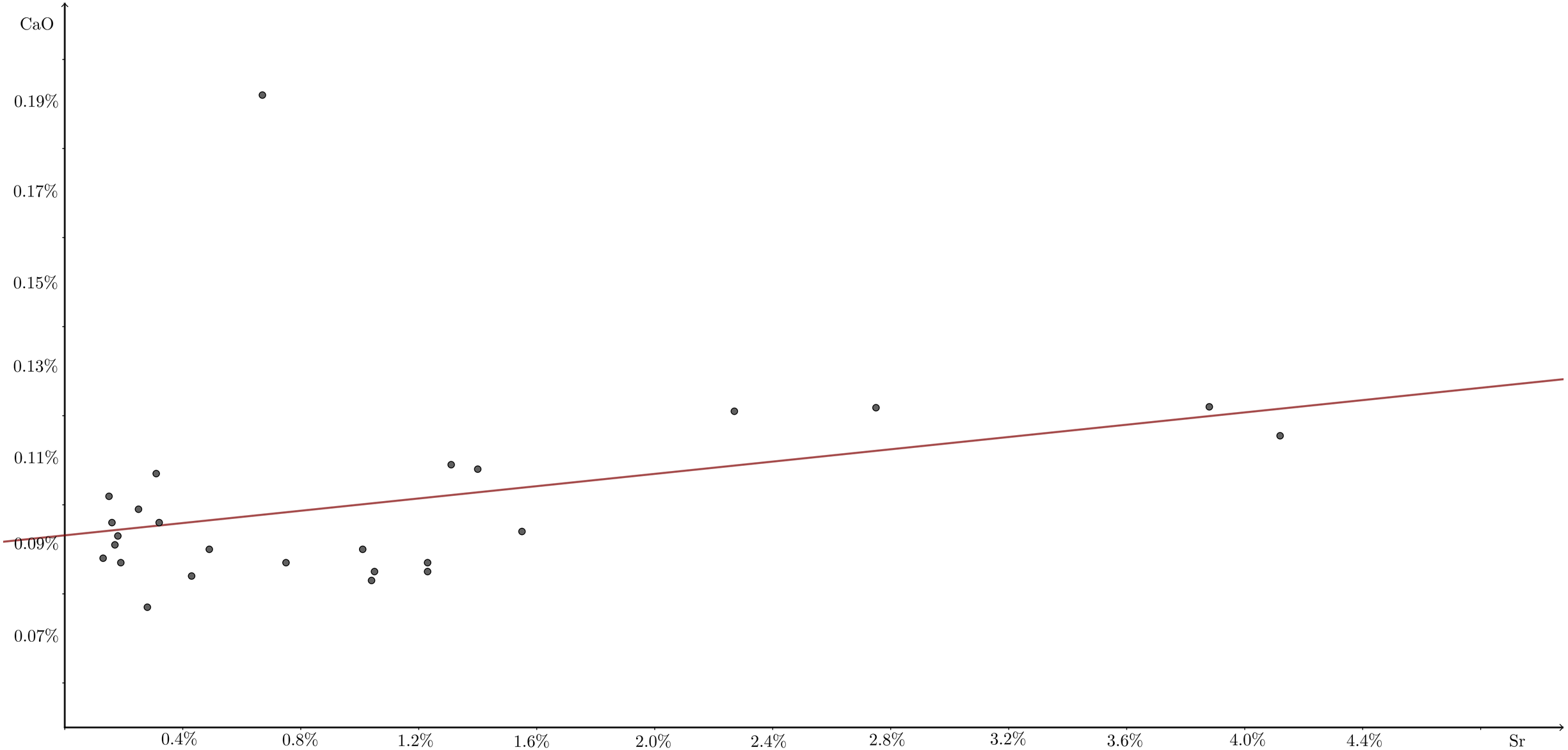

K_2O
4.4%
4.2%
4.0%
3.8%
3.6%
3.4%
3.2%
73%
74%
75%
76%
77%
78%
79%
80%
81%
82%
83%
84%
85%
86%
SiO_2

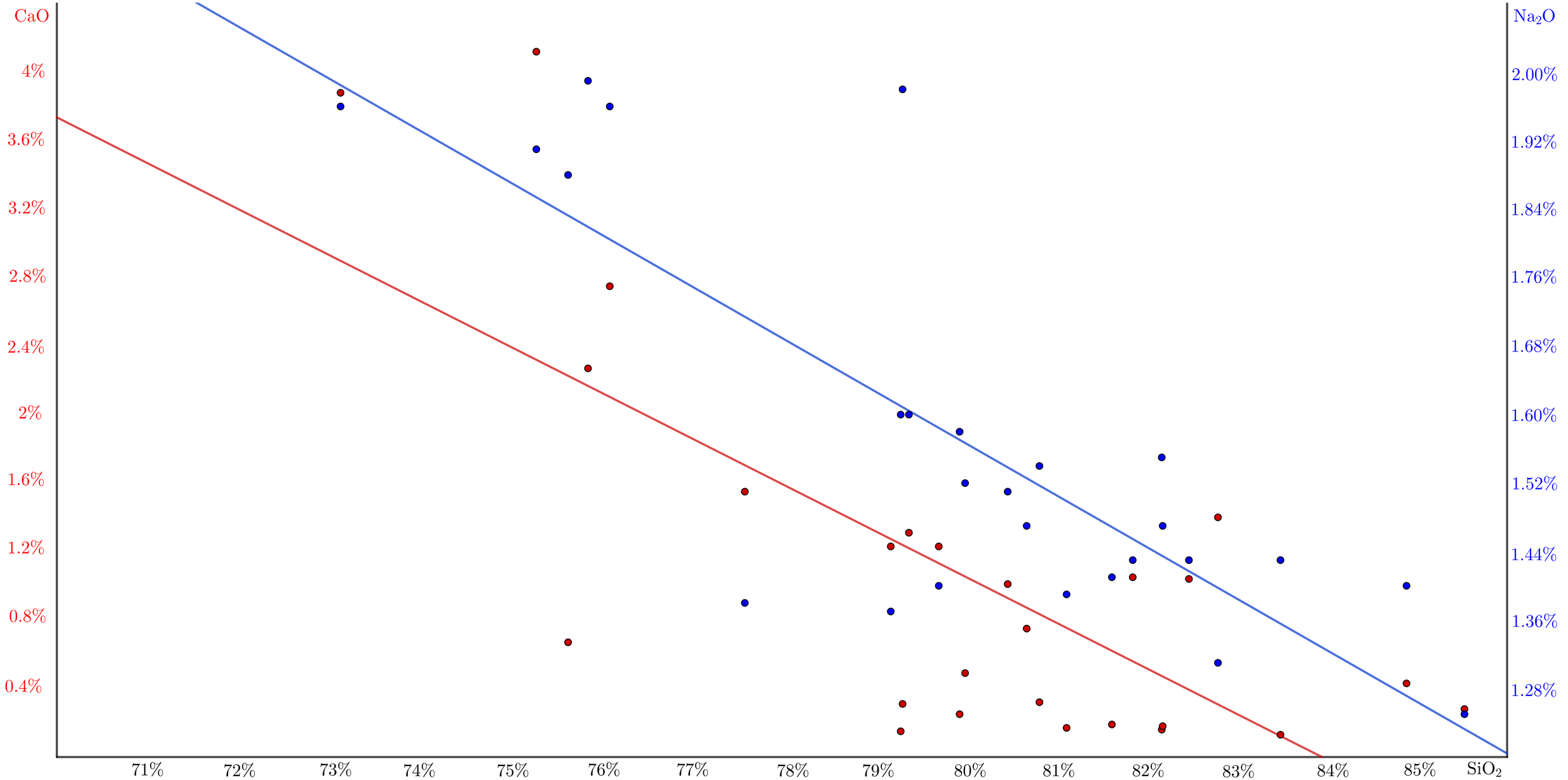

Zr
240
210
180
150
120
90
60
73%
74%
75%
76%
77%
78%
79%
80%
81%
82%
83%
84%
85%
86%
SiO_2

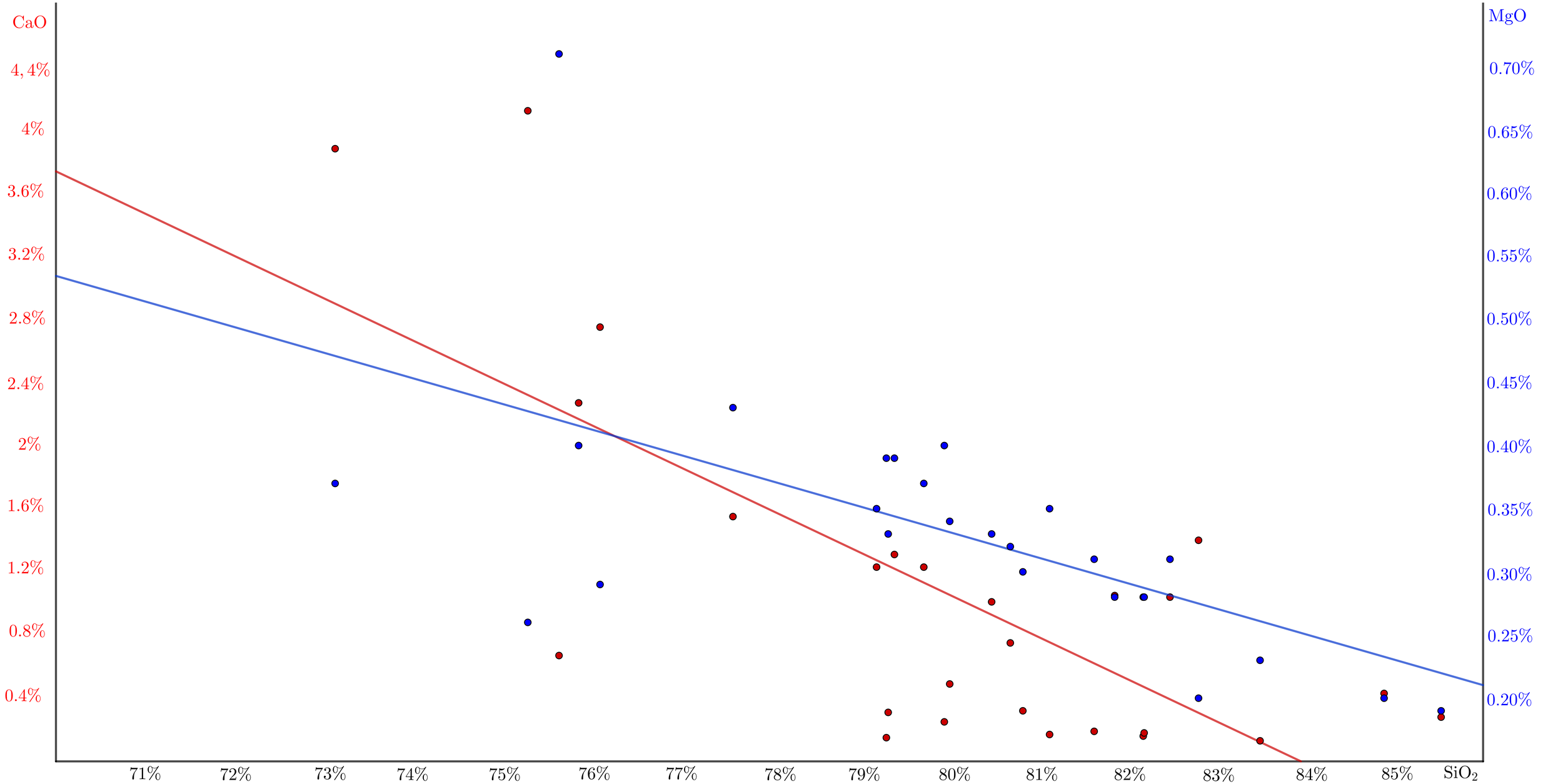

Rb
160
150
140
130
120
110
100
80
90
100
110
120
130
140
150
160
170
180
190
Sr

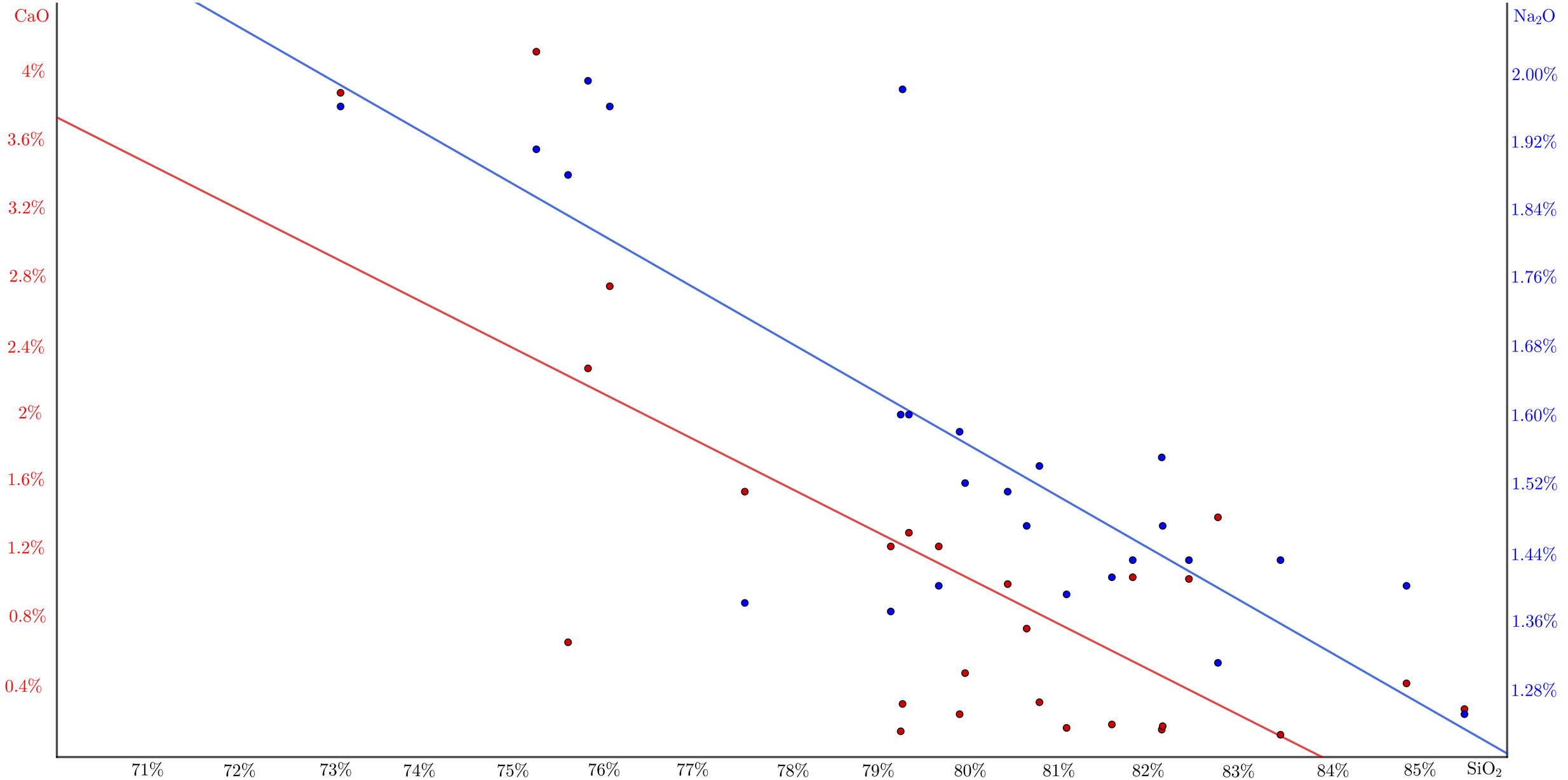

Ba
2100
1800
1500
1200
900
600
80
90
100
110
120
130
140
150
160
170
180
190
Sr

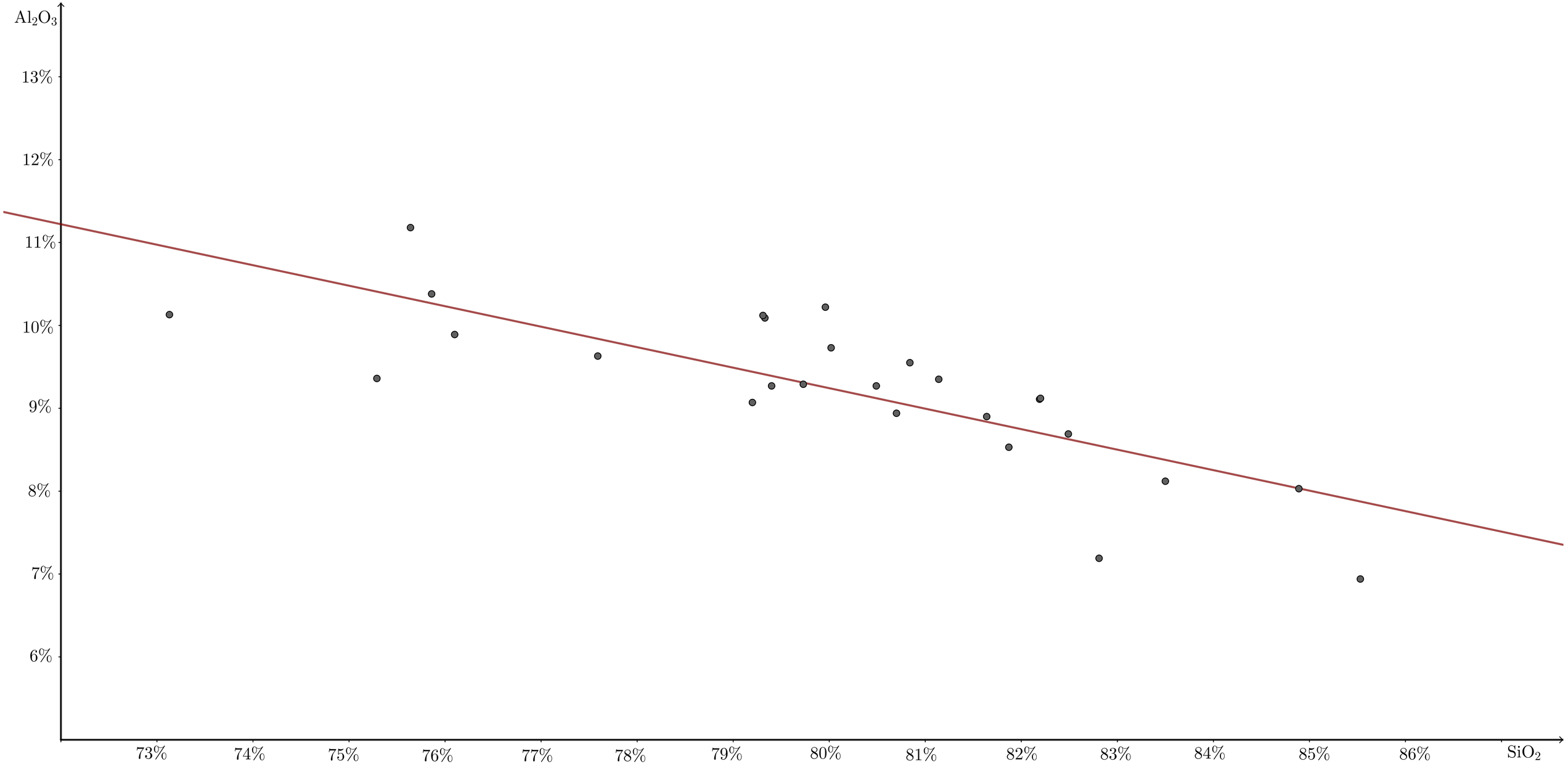

Rb
150
140
130
120
110
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
Ba


CaO
0.19%
0.17%
0.15%
0.13%
0.11%
0.09%
0.07%
0.4%
0.8%
1.2%
1.6%
2.0%
2.4%
2.8%
3.2%
3.6%
4.0%
4.4%
Sr


CaO
4%
3.6%
3.2%
2.8%
2.4%
2%
1.6%
1.2%
0.8%
0.4%
Na2O
2.00%
1.92%
1.84%
1.76%
1.68%
1.60%
1.52%
1.44%
1.36%
1.28%
71%
72%
73%
74%
75%
76%
77%
78%
79%
80%
81%
82%
83%
84%
85%
SiO2


CaO
4, 4%
4%
3.6%
3.2%
2.8%
2.4%
2%
1.6%
1.2%
0.8%
0.4%
MgO
0.70%
0.65%
0.60%
0.55%
0.50%
0.45%
0.40%
0.35%
0.30%
0.25%
0.20%
71%
72%
73%
74%
75%
76%
77%
78%
79%
80%
81%
82%
83%
84%
85%
SiO2


CaO
4%
3.6%
3.2%
2.8%
2.4%
2%
1.6%
1.2%
0.8%
0.4%
Na2O
2.00%
1.92%
1.84%
1.76%
1.68%
1.60%
1.52%
1.44%
1.36%
1.28%
71%
72%
73%
74%
75%
76%
77%
78%
79%
80%
81%
82%
83%
84%
85%
SiO2


Al_2O_3
13%
12%
11%
10%
9%
8%
7%
6%
73%
74%
75%
76%
77%
78%
79%
80%
81%
82%
83%
84%
85%
86%
SiO_2

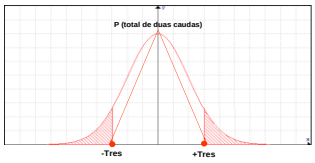
ANEXOS

- Tabela t: distribuição t – student (usada para realização dos testes de hipótese do tipo t de student).

- Planilha contendo banco de dados geoquímicos para o Alogrupo Guaritas. Análises de elementos maiores, menores e traço em rocha total por espectrometria de massa por fusão da amostra com plasma acoplado indutivelmente (FUS-ICP: Fusion-Inductively Coupled Plasma). As análises químicas em rocha total para elementos maiores, menores e traço foram preparadas no Activation Laboratories Ltd. (Laboratório *Actlabs*) sediado no Canadá.

Tabela T: Distribuição de t-Student segundo os graus de liberdade e uma dada probabilidade num teste bicaudal (primeira linha)

Para um teste monocaudal, considere metade do valor de probabilidade apontado



N° de graus de liberdade	Probabilidade para um teste bicaudal													
	0,95	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0787	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,657	636,619
2	0,0708	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,0681	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,0667	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,0659	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,0654	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,0650	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,0647	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,0645	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,0643	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5869
11	0,0642	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370
12	0,0640	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178
13	0,0639	0,1281	0,2586	0,3940	0,5375	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2208
14	0,0638	0,1280	0,2582	0,3933	0,5366	0,6924	0,8681	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	4,1405
15	0,0638	0,1278	0,2579	0,3928	0,5357	0,6912	0,8662	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467	4,0728
16	0,0637	0,1277	0,2576	0,3923	0,5350	0,6901	0,8647	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	4,0150
17	0,0636	0,1276	0,2573	0,3919	0,5344	0,6892	0,8633	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
18	0,0636	0,1274	0,2571	0,3915	0,5338	0,6884	0,8620	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,9216
19	0,0635	0,1274	0,2569	0,3912	0,5333	0,6876	0,8610	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,8834
20	0,0635	0,1273	0,2567	0,3909	0,5329	0,6870	0,8600	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,8495
21	0,0635	0,1272	0,2566	0,3906	0,5325	0,6864	0,8591	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,8193
22	0,0634	0,1271	0,2564	0,3904	0,5321	0,6858	0,8583	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,7921
23	0,0634	0,1271	0,2563	0,3902	0,5317	0,6853	0,8575	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
24	0,0634	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,7454
25	0,0633	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
26	0,0633	0,1269	0,2560	0,3896	0,5309	0,6840	0,8557	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,7066
27	0,0633	0,1268	0,2559	0,3894	0,5306	0,6837	0,8551	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,6896
28	0,0633	0,1268	0,2558	0,3893	0,5304	0,6834	0,8546	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739
29	0,0633	0,1268	0,2557	0,3892	0,5302	0,6830	0,8542	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,6594
30	0,0632	0,1267	0,2556	0,3890	0,5300	0,6828	0,8538	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,6460
60	0,0630	0,1262	0,2545	0,3872	0,5272	0,6786	0,8477	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	3,4602
90	0,0629	0,1260	0,2541	0,3866	0,5263	0,6772	0,8456	1,0424	1,2910	1,6620	1,9867	2,3685	2,6316	3,4019
120	0,0628	0,1259	0,2539	0,3862	0,5258	0,6765	0,8446	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174	3,3735
150	0,0628	0,1259	0,2538	0,3861	0,5255	0,6761	0,8440	1,0400	1,2872	1,6551	1,9759	2,3515	2,6090	3,3566
180	0,0628	0,1258	0,2537	0,3859	0,5253	0,6759	0,8436	1,0394	1,2863	1,6534	1,9732	2,3472	2,6034	3,3454
210	0,0628	0,1258	0,2537	0,3858	0,5252	0,6757	0,8433	1,0390	1,2856	1,6521	1,9713	2,3442	2,5994	3,3375
240	0,0628	0,1258	0,2536	0,3858	0,5251	0,6755	0,8431	1,0387	1,2851	1,6512	1,9699	2,3420	2,5965	3,3315
270	0,0628	0,1258	0,2536	0,3857	0,5250	0,6754	0,8430	1,0384	1,2847	1,6505	1,9688	2,3402	2,5942	3,3269
300	0,0628	0,1258	0,2536	0,3857	0,5250	0,6753	0,8428	1,0382	1,2844	1,6499	1,9679	2,3388	2,5923	3,3233
400	0,0627	0,1257	0,2535	0,3856	0,5248	0,6751	0,8425	1,0378	1,2837	1,6487	1,9659	2,3357	2,5882	3,3150
500	0,0627	0,1257	0,2535	0,3855	0,5247	0,6750	0,8423	1,0375	1,2832	1,6479	1,9647	2,3338	2,5857	3,3101
800	0,0627	0,1257	0,2534	0,3855	0,5246	0,6748	0,8421	1,0371	1,2826	1,6468	1,9629	2,3310	2,5820	3,3027
1000	0,0627	0,1257	0,2534	0,3854	0,5246	0,6747	0,8420	1,0370	1,2824	1,6464	1,9623	2,3301	2,5808	3,3003

Final Report
Activation Laboratories

Analyte Symbol	SiO2	Al2O3	Fe2O3(T)	MnO	MgO	CaO	Na2O	K2O	TiO2	P2O5	LOI	Total	Sc	Be	V
Unit Symbol	%	%	%	%	%	%	%	%	%	%	%	%	ppm	ppm	ppm
Detection Limit	0.01	0.01	0.01	0.001	0.01	0.01	0.01	0.01	0.001	0.01		0.01	1	1	5
Analysis Method	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP
GXR-1 Meas															
GXR-1 Cert															
WMG-1 Meas															
WMG-1 Cert															
NIST 694 Meas	11,5	1,88	0,72	0,012	0,32	43,45	0,85	0,57	0,112	30,15					1659
NIST 694 Cert	11,2	1,8	0,79	0,012	0,33	43,6	0,86	0,51	0,11	30,2					1737
DNC-1 Meas	47,16	18,4	10	0,147	10,19	11,29	1,91	0,2	0,492	0,08			31	< 1	154
DNC-1 Cert	47,04	18,3	9,93	0,149	10,05	11,27	1,87	0,23	0,48	0,09			31	1	148
BIR-1 Meas	48,11	15,63	11,52	0,171	9,65	13,26	1,84	0,06	0,972	0,02			43	< 1	335
BIR-1 Cert	47,77	15,35	11,26	0,171	9,68	13,24	1,75	0,03	0,96	0,05			44	1	313
GXR-2 Meas															
GXR-2 Cert															
FK-N Meas	64,95	18,45	0,1	0,002	< 0.01	0,1	2,48	12,79	0,007	0,01			< 1	< 1	6
FK-N Cert	65,02	18,61	0,09	0,005	0,01	0,11	2,58	12,81	0,02	0,02			1	1	5
LKSD-3 Meas															
LKSD-3 Cert															
MAG-1 (Depleted) Meas															
MAG-1 (Depleted) Cert															
NIST 1633b Meas	49,03	28,34	11,41	0,017	0,78	2,14	0,25	2,3	1,304	0,57			41		302
NIST 1633b Cert	49,24	28,43	11,13	0,02	0,8	2,11	0,27	2,35	1,32	0,53			41		296
SY-2 Meas	59,55	11,81	6,32	0,32	2,63	7,99	4,21	4,47	0,136	0,45			7	21	77
SY-2 Cert	60,1	12,12	6,27	0,32	2,7	7,98	4,34	4,48	0,14	0,43			7	23	52
W-2a Meas	52,76	15,38	11,03	0,171	6,37	10,94	2,24	0,61	1,082	0,14			36	< 1	277
W-2a Cert	52,44	15,35	10,74	0,163	6,37	10,87	2,14	0,63	1,06	0,13			36	1	262
NIST 696 Meas	3,91	52,95	8,47	0,004	0,01	0,02		0,04	2,57	0,05					395
NIST 696 Cert	3,79	54,5	8,7	0,004	0,01	0,02		0,01	2,64	0,05					403
JSD-3 Meas	76,41	9,9	4,32	0,145	1,16	0,56	0,4	2	0,414	0,09					
JSD-3 Cert	76	9,91	4,37	0,148	1,17	0,56	0,41	1,97	0,403	0,08					
G15 Orig	80,29	9,26	1,11	0,014	0,34	1,01	1,51	3,72	0,155	0,07	1,91	99,38	1	2	13
G15 Dup	80,68	9,28	1,46	0,014	0,33	1	1,51	3,73	0,152	0,06	1,91	100,1	1	2	12
G23 Orig	79,33	10,09	1,7	0,06	0,33	0,31	1,98	3,56	0,201	0,06	1,24	98,86	2	2	13
G23 Split	79,31	10,13	1,72	0,06	0,33	0,31	1,99	3,62	0,197	0,06	1,21	98,93	2	2	16
Method Blank Method Blank															

Ba	Sr	Y	Zr	Cr	Co	Ni	Cu	Zn	Ga	Ge	As	Rb	Nb	Mo	Ag	In	Sn
ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm
3	2	2	4	20	1	20	10	30	1	1	5	2	1	2	0.5	0.2	1
FUS-ICP	FUS-ICP	FUS-ICP	FUS-ICP	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS
				< 20	8	30	1110	740	16		425	4	1	18	33,7	0,8	54
				12	8	41	1110	760	14		427	14	1	18	31	0,8	54
				780	207	2750	5270	130	10		15		5	< 2	0,9		2
				770	200	2700	5900	110	10		7		6	1	2,7		2
107	143	17	37	280	55	250	110	90	14	1	< 5	4	1	< 2	< 0.5		
114	145	18	41	285	55	247	96	66	15	1	0	5	3	1	0		
7	110	14	13	390	51	170	140	110	15	1	< 5	< 2	< 1	< 2	< 0.5		< 1
7	108	16	16	382	51	166	126	71	16	2	0	0	1	1	0		1
				40	9	< 20	110	550	36		37	80	11	< 2	12,4	< 0.2	2
				36	9	21	76	530	37		25	78	11	2	17	0,3	2
208	38	< 2	< 4														
200	39	1	13														
				80	32	40	40	170			51	76	7	< 2	< 0.5		2
				87	30	47	35	152			27	78	8	2	2,7		3
				100	22	40	30	170	22		18	142	12	< 2	< 0.5	< 0.2	4
				97	20	53	30	130	20		9	149	12	2	0,1	0,2	4
712	1039																
709	1041																
456	271	135	289														
460	275	130	280														
175	198	21	90	90	43	70	120	120	17	2	< 5	20	7	< 2	< 0.5		
182	190	24	94	92	43	70	110	80	17	1	1	21	8	1	0		
			1053														
			1037														
562	90	10	132	100	1	< 20	10	40	9	2	11	134	5	< 2	< 0.5	< 0.2	1
567	90	10	134	100	1	< 20	10	30	9	2	13	128	5	< 2	< 0.5	< 0.2	1
900	107	14	126	140	2	< 20	10	40	11	2	13	133	7	< 2	< 0.5	< 0.2	2
902	108	14	120	130	2	< 20	< 10	40	11	2	14	131	7	< 2	< 0.5	< 0.2	2
				< 20	< 1	< 20	< 10	< 30	< 1	< 1	< 5	< 2	< 1	< 2	< 0.5	< 0.2	< 1

Sb	Cs	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Hf	Ta	W
ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm
0.5	0.5	0.1	0.1	0.05	0.1	0.1	0.05	0.1	0.1	0.1	0.1	0.1	0.05	0.1	0.04	0.2	0.1	1
FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS
122	2,8	8,3	15,9		8,6	3	0,72	3,8	0,8	4,9			0,43	2,2	0,31	1	< 0.1	164
122	3	7,5	17		18	2,7	0,69	4,2	0,8	4,3			0,43	1,9	0,28	1	0,2	164
4,9	0,5	8,7	17,6		9,4	2,4	0,76		0,4	2,5	0,5		0,22	1,3	0,19	1,6	0,3	< 1
1,8	0,5	8,2	16		9	2,3	0,82		0,3	2,8	0,5		0,2	1,3	0,21	1,3	0,5	1
0,8	< 0.5	4,1	8,8	1,05	4,9	1,4	0,62	1,8	0,4	2,8	0,6	2	0,33	1,9	0,3	1,1	< 0.1	< 1
1	0,3	3,8	10,6	1,3	4,9	1,4	0,59	2	0,4	2,7	0,6	2	0,38	2	0,32	1	0,1	0
< 0.5	< 0.5	0,8	2,2	0,39	2,4	1,1	0,54	1,6	0,4	2,7	0,6	1,7	0,29	1,6	0,25	0,7	< 0.1	< 1
0,6	0	0,6	2	0,38	2,5	1,1	0,54	1,9	0,4	2,5	0,6	1,7	0,26	1,6	0,26	0,6	0	0
30,5	5,5	27,6	54		20,6	3,8	0,9	3,5	0,5	3			0,29	1,8	0,28	7,5	0,9	2
49	5,2	25,6	51,4		19	3,5	0,81	3,3	0,5	3,3			0,3	2	0,27	8,3	0,9	2
0,8	2,4	51,9	93,6		44,9	8,1	1,52		0,9	5,1				2,6	0,41	4,8	0,7	< 1
1,3	2,3	52	90		44	8	1,5		1	4,9				2,7	0,4	4,8	0,7	2
1,2	8,5	42,3	83,2	9,33	36,1	7,1	1,45	6,2	0,9	5,1	0,9	2,7	0,42	2,4	0,36	3,5	1,1	1
1	8,6	43	88	9,3	38	7,5	1,6	5,8	1	5,2	1	3	0,43	2,6	0,4	3,7	1,1	1
< 0.5	0,9	12	25,1		13	3,3	1,15		0,6	3,9	0,8	2,3	0,35	2	0,3	2,7	0,5	< 1
0,8	1	10	23		13	3,3	1		0,6	3,6	0,8	2,5	0,38	2,1	0,33	2,6	0,5	0
1,8	3,3	23,7	36,1	4,51	14,4	2,6	0,54	2,3	0,3	1,9	0,4	1,2	0,18	1,1	0,16	3,9	0,6	< 1
0,8	3,1	22,4	33,8	4,3	13,3	2,5	0,5	2,1	0,3	1,9	0,4	1,2	0,17	1	0,16	3,8	0,5	< 1
< 0.5	3,3	36,8	79	8,74	29,9	5,9	1,17	4,3	0,5	2,7	0,5	1,6	0,21	1,3	0,18	3,7	0,8	< 1
1,1	3,4	36,1	76,8	8,6	29	5,8	1,1	4,1	0,5	2,7	0,5	1,5	0,21	1,3	0,19	3,5	0,8	< 1
< 0.5	< 0.5	< 0.1	< 0.1	< 0.05	< 0.1	< 0.1	< 0.05	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.05	< 0.1	< 0.04	< 0.2	< 0.1	< 1

Tl	Pb	Bi	Th	U
ppm	ppm	ppm	ppm	ppm
0.1	5	0.4	0.1	0.1
FUS-MS	FUS-MS	FUS-MS	FUS-MS	FUS-MS
0,5	730	1380	2,6	31,9
0,4	730	1380	2,4	34,9
	19		1,2	0,7
	15		1,1	0,6
< 0.1	7	< 0.4	0,2	< 0.1
0	6	0	0,2	0,1
< 0.1	< 5	< 0.4	< 0.1	< 0.1
0	3	0	0	0
1	429	< 0.4	8,9	3
1	690	0,7	8,8	2,9
	32		11,5	4,5
	29		11,4	4,6
0,5	30	< 0.4	11,7	2,8
0,6	24	0,3	11,9	2,7
0,1	8	< 0.4	2,3	0,5
0,2	9	0	2,4	0,5
0,8	16	< 0.4	7,9	1,3
0,7	14	< 0.4	7,6	1,2
0,7	16	0,6	10,5	1,3
0,8	26	0,5	10,4	1,3
< 0.1	< 5	< 0.4	< 0.1	< 0.1

ANEXO I
Título da Dissertação/Tese:
"**Abordagem estatística aplicada à análise e interpretação geoquímica do Alogrupo Guaritas, Bacia do Camaquã – RS, Brasil**"
Área de Concentração: Geoquímica
Autor: **Ariane Kravczyck Bernardes**
Orientador: Profa. Dra. Ana Maria Pimentel Mizusaki
Examinador: Rogerio Roque Rubert
Data: 19/11/2021
Conceito: A
PARECER: A presente dissertação apresenta o desenvolvimento de tratamento e abordagem de estatísticos de dados de estratigrafia química, como um experimento como uma ferramenta de suporte a outras técnicas analíticas utilizadas em uma análise de bacias e outros ramos das geociências, visando o aprimoramento de base nos dados geoquímicos com a utilização de dados do Alogrupo Guaritas (Bacia do Camaquã). A metodologia e objetivos estão colocados de forma clara, assim como os resultados mostraram-se consistentes para atender a demonstração da efetividade da ferramenta tanto no presente caso como em casos onde é necessária a otimização do uso de dados e de recursos em projetos de enfoque e finalidades variadas. Como aspectos a serem observados, algumas informações poderiam estar melhor definidas no volume tais como critérios de seleção de pontos de amostragem, indicação dos níveis amostrados, melhor apresentação de algumas figuras e gráficos e formatação de alguns intervalos de texto. O trabalho atingiu o objetivo com êxito estando apto para aprovação.
Assinatura: Data:18/11/2021
Ciente do Orientador: Ciente do Aluno:

ANEXO I
Título da Dissertação/Tese:
"**Abordagem estatística aplicada à análise e interpretação geoquímica do Alogrupo Guaritas, Bacia do Camaquã – RS, Brasil**"
Área de Concentração: Geoquímica
Autor: **Ariane Kravczyck Bernardes**
Orientador: Profa. Dra. Ana Maria Pimentel Mizusaki
Examinador: Prof. Dr. Giovani Matte Cioccari
Data: 17/11/2021
Conceito: B
PARECER: A dissertação de mestrado da aluna Ariane Kravczyck Bernades possui um texto bem estruturado, de fácil leitura, sem muitos erros semânticos, entretanto apresenta alguns erros de escrita em palavras como por exemplo: sem itálico em palavras estrangeiras, abreviações sem pontuação, falta de acentuação em algumas palavras, além de desestruturação do texto e figuras em algumas páginas. O título define bem o trabalho. O resumo e o abstract apresentam problemas de estruturação e defini mal a proposta do trabalho. A introdução é sucinta e bem estruturada, contudo carece de maior relação com o tema proposto na dissertação. A justificativa é longa e confusa, contudo identifica-se a proposta de um problema e sua resolução através da abordagem estatística utilizando dados de estratigrafia química e geoquímica. Os objetivos principal e secundários estão claros, contendo metas bem definidas. O capítulo de "Área de Estudo e Dados Analisados" apresenta uma discussão sucinta sobre a Bacia do Camaquã do ponto de vista tectonossedimentar e datação das unidades em estudo, apresentando alguns equívocos na numeração das figuras. Os capítulos 4 (Estratigrafia Química) e 5 (Geoestatística) foram escritos de maneira muito sucinta, cabendo a autora dissertar melhor sobre os conceitos, as técnicas, as ferramentas, os métodos, etc. A metodologia é bem descrita, abordando as diferentes técnicas analíticas para resolução do problema proposto. Os resultados são apresentados em forma de artigo científico, como normas estabelecidas pelo Programa de Pós-Graduação em Geociências da UFRGS. Os resultados e discussões apresentam problemas de configuração dos parágrafos, não citação das figuras no texto e os resultados não estão claros no corpo do texto. As discussões dos resultados apresentam dados consistem com o Aloformações Pedra Pintada e Varzinha do Alogrupo Guaritas da Bacia do Camaquã, mostrando-se bem embasadas teoricamente com levantamentos geoquímicos robustos, entretanto faltou discutir as análises de correlação linear e ACP e compará-los com dados de bibliografia, para fortalecer as evidências observadas no trabalho. As conclusões e considerações finais demostram o estabelecimento de uma significância estatística sobre quais composições químicas são mais distintivas para as regiões de amostradas ao longo do perfil colunar no

Alogrupo Guaritas. A bibliografia utilizada é atual e dentro da área do estudo, com trabalhos clássicos de sobre a geologia da Bacia do Camaquã, bem como interpretações estatísticas, geoquímicas e geocronológicas. De maneira geral, a contextualização do problema está clara e a solução apresentada na dissertação estão ótimos, no entanto os objetivos diz que estudo desenvolverá um fluxo de etapas necessárias para o tratamento dos dados, que não foi demonstrado claramente durante os resultados e conclusões do trabalho. Os resultados mostram que os objetivos foram alcançados e que a técnica apresentada funcionou de maneira satisfatória para as rochas sedimentares ordovicianas da Bacia do Camaquã.

Assinatura:	Data: 17/11/2021
Ciente do Orientador: Ciente do Aluno:	

ANEXO I
Título da Dissertação/Tese:
"**Abordagem estatística aplicada à análise e interpretação geoquímica do Alogrupo Guaritas, Bacia do Camaquã – RS, Brasil**"
Área de Concentração: Geoquímica
Autor: **Ariane Kravczyck Bernardes**
Orientador: Profa. Dra. Ana Maria Pimentel Mizusaki
Examinador: Cassiana Roberta Lizzoni Michelin
Data: 18 de novembro de 2021
Conceito: B
PARECER: A discente Ariane Kravczyck Bernardes desenvolveu a dissertação intitulada "*Abordagem estatística aplicada à análise e interpretação geoquímica do Alogrupo Guaritas, Bacia do Camaquã – RS, Brasil*" de forma clara e satisfatória. O tema escolhido é inovador e representa uma ferramenta relevante à análise de dados químicos, a elaboração de um fluxo de etapas necessárias para o tratamento dos dados e aplicação das técnicas estatísticas selecionadas auxilia grandemente na interpretação de dados geoquímicos de rochas sedimentares. Como produto da dissertação, orientanda e orientadora submeteram o artigo científico intitulado "*Statistical methods applied to geochemical analysis of data from the Guaritas Allogroup, Camaquã Basin – RS, Brazil*."O manuscrito está bem estruturado e redigido, e apresenta resultados estatísticos bem fundamentados. Sente-se necessidade de uma correlação mais intensa com os dados geoquímicos, buscando-se assim um aprofundamento nas interpretações geológicas que podem ser realizadas utilizando-se a metodologia proposta. Por fim, parabenizo a orientanda e orientadora pela qualidade do trabalho apresentado.
Assinatura: Data: 18 de novembro de 2021
Ciente do Orientador: Ciente do Aluno: