

**THIAGO LUIS ALVES JATOBÁ**

**ANÁLISE POR AGRUPAMENTO APLICADA AO TRATAMENTO DE  
MINÉRIOS**

Trabalho de Formatura em Engenharia de  
Minas do curso de graduação do Departamento  
de Engenharia de Minas e de Petróleo da  
Escola Politécnica da Universidade de São  
Paulo.

Orientação: Prof. Dr. Homero Delboni Jr.

**São Paulo  
2012**

**THIAGO LUIS ALVES JATOBÁ**

**ANÁLISE POR AGRUPAMENTO APLICADA AO TRATAMENTO DE  
MINÉRIOS**

**São Paulo  
2012**

TF-2012

J318a

Syno 2448816

H2012f

DEDALUS - Acervo - EPMI



31700009736

#### FICHA CATALOGRÁFICA

**Jatobá, Thiago Luis Alves**

**Análise por agrupamento aplicada ao tratamento de minérios**

**/ T.L.A. Jatobá. -- São Paulo, 2012.**

**49 p.**

**Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Minas e de Petróleo.**

**1. Minérios (Tratamento) 2. Análise de conglomerados I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Minas e de Petróleo II. t.**

## **FICHA CATALOGRÁFICA**

## AGRADECIMENTOS

Ao Professor Dr. Homero Delboni Jr., após todos estes anos, não somente pela inspiração, pelo incentivo e pela confiança, sem os quais teria sido impossível prosseguir nesta jornada, mas em incontáveis lições, ministradas dentro e fora das salas de aula.

À Professora Dr<sup>a</sup>. Ana Carolina Chierigati, por dispor de seu precioso tempo e vasto conhecimento, engrandecendo de forma irreversível as tarefas de pesquisa pretendidas por mim.

Aos amigos Wellington Gomes, Bianca Foggiatto, Gustavo Angelocci, Célio Vitor Marques e Thammiris El Hajj, por dividirem não só seu valioso conhecimento, mas também aprazíveis momentos de trabalho e lazer.

Aos técnicos Juninho, Marcos e Rogério por terem sido sempre solícitos e competentes, nas atividades laboratoriais e fora dela

Aos meus pais e a minha irmã por todo apoio dedicado, esforços realizados, orientações bem intencionadas, cumplicidade e paciência durante anos.

À minha esposa, por estar sempre ao meu lado, de mãos dadas; à frente, guiando-me; ou atrás, impulsionando-me. A ela, pelo imenso prazer de construir, em conjunto, este intenso e gratificante caminho, brita por brita.

“É tão absurdo dizer que um homem não pode  
amar a mesma mulher toda a vida, quanto dizer  
que um violinista precisa de diversos violinos  
para tocar a mesma música.”

(Honoré de Balzac)

## RESUMO

Foram apresentadas técnicas de análise por agrupamento, aplicadas ao tratamento de 50 amostras de minério de ferro de um empreendimento em operação. Ensaios padronizados de separação em líquido denso e flotação foram conduzidos e, os teores de ferro dos concentrados, utilizados no estudo de similaridade entre as amostras. Dois grupos de desempenhos distintos foram identificados utilizando-se o método de Ward, como técnica de análise por agrupamento hierárquico, e o método *K-means*, como técnica de análise por agrupamento particional. As distinções entre os grupos foram validadas estatisticamente através de testes de comparação entre as médias dos teores de ferro.

**Palavras-chave:** Tratamento de minérios, análise por agrupamento.

## ABSTRACT

Cluster analysis procedures were presented and applied to data processing of 50 samples of iron ore. Heavy liquid separation and froth flotation standardized tests were conducted and iron grades were used to assess similarity between samples. Two groups of different behaviours were identified using the Ward's method, as a criterion of hierarchical cluster analysis, and the *K-means* method, as a technique for partitional cluster analysis. Groups contrasts were statistically validated through comparison tests between the iron grades means.

**Keywords:** Mineral processing, cluster analysis.

## LISTA DE FIGURAS

|                                                                                |    |
|--------------------------------------------------------------------------------|----|
| Figura 1 - Métodos de agrupamento hierárquico.....                             | 18 |
| Figura 2 - Representação esquemática do método de agrupamento particional..... | 21 |
| Figura 3 - Fotografias dos ensaios de separação em líquido denso.....          | 26 |
| Figura 4 - Fotografias dos ensaios de deslamagem. ....                         | 27 |
| Figura 5 - Fluxograma dos estágios dos ensaios de flotação.....                | 28 |
| Figura 6 - Máquina marca Denver empregada nos ensaios de flotação. ....        | 28 |

## LISTA DE TABELAS

|                                                                                  |    |
|----------------------------------------------------------------------------------|----|
| Tabela 1 - Condições estabelecidas para os ensaios de flotação. ....             | 29 |
| Tabela 2 - Teores de ferro nos concentrados. ....                                | 33 |
| Tabela 3 - Análise descritiva das variáveis C1, C2 e C3. ....                    | 34 |
| Tabela 4 - Testes de normalidade. ....                                           | 34 |
| Tabela 5 - Matriz de correlação de Spearman. ....                                | 35 |
| Tabela 6 - Composição dos grupos e distâncias dos elementos aos centróides. .... | 36 |
| Tabela 7 - Análise descritiva por grupo das variáveis C1, C2 e C3. ....          | 37 |
| Tabela 8 - Testes de normalidade. ....                                           | 37 |
| Tabela 9 - Comparação entre grupos. ....                                         | 38 |

# SUMÁRIO

|                                                             |           |
|-------------------------------------------------------------|-----------|
| <b>1 INTRODUÇÃO .....</b>                                   | <b>9</b>  |
| <b>2 OBJETIVOS .....</b>                                    | <b>12</b> |
| <b>3 REVISÃO BIBLIOGRÁFICA .....</b>                        | <b>13</b> |
| <b>3.1 Definição do conceito.....</b>                       | <b>13</b> |
| <b>3.2 Aplicações.....</b>                                  | <b>14</b> |
| <b>3.3 Técnicas de agrupamento .....</b>                    | <b>15</b> |
| <b>3.4 Métodos de Associação.....</b>                       | <b>16</b> |
| <b>3.4.1 Agrupamento Hierárquico Aglomerativo .....</b>     | <b>17</b> |
| <b>3.4.1.1 Single-Linkage.....</b>                          | <b>18</b> |
| <b>3.4.1.2 Complete-Linkage .....</b>                       | <b>19</b> |
| <b>3.4.1.3 Average-Linkage .....</b>                        | <b>19</b> |
| <b>3.4.1.4 Método de Ward.....</b>                          | <b>20</b> |
| <b>3.4.2 Agrupamento Particional - <i>K-Means</i> .....</b> | <b>20</b> |
| <b>3.5 Medidas de similaridade.....</b>                     | <b>22</b> |
| <b>4 MATERIAIS E MÉTODOS.....</b>                           | <b>24</b> |
| <b>4.1 Análises Químicas.....</b>                           | <b>24</b> |
| <b>4.2 Peneiramento .....</b>                               | <b>24</b> |
| <b>4.3 Separação em Líquido Denso .....</b>                 | <b>25</b> |
| <b>4.4 Deslamagem .....</b>                                 | <b>26</b> |
| <b>4.5 Flotação .....</b>                                   | <b>27</b> |
| <b>4.6 Agrupamento de Amostras.....</b>                     | <b>30</b> |
| <b>5 RESULTADOS E DISCUSSÕES .....</b>                      | <b>32</b> |
| <b>6 CONCLUSÕES.....</b>                                    | <b>40</b> |
| <b>7 REFERÊNCIAS .....</b>                                  | <b>41</b> |
| <b>APÊNDICE A.- Gráficos.....</b>                           | <b>44</b> |

## 1 INTRODUÇÃO

A indústria mineral possui particularidades que a diferem de outros empreendimentos industriais. Segundo Chaves (1988), um dos aspectos intrínsecos à mineração é a unicidade das características geológicas, topográficas e geográficas de cada jazimento, tornando indispensáveis análises aprofundadas e individualizadas do material pesquisado. Cada novo projeto pressupõe um estudo de um processo de beneficiamento único, mesmo que haja convergências com outros empreendimentos instaurados. Unem-se a isto grandes prazos de maturação do empreendimento, que permeiam horizontes de oito a dez anos, período este que se inicia com a constatação de anomalias geológicas (através de pesquisas geofísicas) e caminha paulatinamente na direção de incrementar gastos com pesquisas que tragam informações para estudos de viabilidade econômica, permitindo aos portadores de recursos tomarem decisões e então arcarem com os elevados riscos dos projetos de implantação de indústrias minerais.

Ainda, indústrias de transformação usualmente lidam com limites relativamente estreitos da variabilidade de seus insumos, o que não acontece com a mineração, que tem de conviver com a dificuldade de parametrização de seus minérios, oriundos da diversidade da natureza.

A indústria mineira, no que tange a aspectos tecnológicos, pode ser grosseiramente segmentada em duas especificidades, quais sejam, a lavra, comprometida com a logística e com a execução de extração do bem mineral, e o beneficiamento do insumo lavrado, que submete o produto da lavra a processos físicos e químicos, objetivando suprir demandas de matérias primas de outros setores industriais de transformação, como exemplos, o metalúrgico e as empresas consumidoras de agregados para construção civil.

Na implantação de um projeto mineiro, a determinação de rotas de beneficiamento que permitam alcançar as especificações de um determinado produto, constitui-se de processos investigatórios gradativos sobre a resposta tecnológica do material quando submetido a ensaios, geralmente, consagrados. Estes são inicialmente realizados em escalas mássicas reduzidas, por laboratórios capacitados e, obtendo-se resultados promissores,

permitem progressivos aumentos nos investimentos, que visam elevar a representatividade das amostragens e os níveis de previsibilidade de desempenho tecnológico do material estudado.

Assim, é pertinente que tais estudos de desempenho em escalas de bancada sejam otimizados de forma a produzir o máximo de informações úteis no processo decisório das empresas.

Ainda, instalações em operação enfrentam frequentes dificuldades na obtenção de produtos especificados conforme novas porções do sítio mineral sejam incorporadas na lavra e, conseqüentemente, na alimentação da planta de processamento mineral. Etapas de expansão de produção também demandam estudos em escalas de bancada ou em escalas piloto, estes últimos, ensaios contínuos de magnitude reduzida, porém mais representativos do que os ensaios de batelada realizados em laboratórios. Analogamente, empresas que desejem sanar entraves em seus processos, usualmente recorrem a ensaios laboratoriais.

Nesse contexto de investimentos cautelosos, os gastos de recursos com pesquisas subsidiadoras de informações devem ser distribuídos estrategicamente, de forma a direcionar o capital para avaliações técnicas e mercadológicas condizentes com o risco inerente a cada etapa do processo de avaliação do empreendimento.

Portanto, a otimização dos recursos alocados em definições de rotas de beneficiamento é valiosa para o empreendimento, e não deve ser suprimida, nem pela ausência de representatividade das amostras coletadas, nem pela falta de aprofundamento nos ensaios de caracterização tecnológica.

Tendo em vista a alta heterogeneidade imposta pela natureza aos corpos de minérios, salvo raras exceções, os números de amostras necessárias para a obtenção de análises representativas das jazidas são elevados. A realização de numerosos testes sobre as amostras torna o processo lento e dispendioso.

O presente trabalho apresenta técnicas multivariadas de análise por agrupamento, que possibilitam, tanto a redução do número de ensaios, quanto o aprofundamento da previsibilidade do comportamento tecnológico do minério, através da realização de ensaios preliminares individualizados, que subsidiem a geração dos parâmetros de desempenho

requeridos. Nesta abordagem, os ensaios individualizados devem ser padronizados de modo que os parâmetros de interesse (Ex.: teores, recuperações, índices de quebra, *bond work index*, índices de seletividade) sejam comparáveis após a realização dos procedimentos. De posse dos resultados, procede-se à análise exploratória dos dados (AED), utilizando-se ferramentas da estatística descritiva e aplicação de técnicas de agrupamento.

Após a validação do procedimento, realiza-se criteriosa homogeneização física das amostras em grupos, que são, então, conduzidos a etapas de aprofundamento da caracterização tecnológica e desenvolvimento de rotas de beneficiamento, de forma que o comportamento de cada amostra seja satisfatoriamente representado pelos resultados obtidos para o grupo ao qual ela foi enquadrada.

O conceito descrito anteriormente pode ser sumarizado nas seguintes etapas:

1) Definição do problema:

- Determinação do protocolo de amostragem e extração das amostras;
- Definição dos parâmetros de interesse;
- Determinação da rota de ensaios necessária;
- Avaliação do conjunto de dados.

2) Determinação do conjunto de técnicas a serem aplicadas:

- Seleção dos casos, parâmetros e métodos a serem empregados na análise de agrupamento;
- Aplicação dos métodos;
- Avaliação dos resultados;
- Validação da técnica.

3) Aprofundamento dos estudos de caracterização:

- Homogeneização das massas relativas às amostras de cada grupo;
- Determinação de rotas de ensaios de aprofundamento;
- Realização dos ensaios;
- Análise dos resultados.

## **2 OBJETIVOS**

O objetivo do presente trabalho é apresentar técnicas de análise por agrupamento, conforme aplicadas a um conjunto de dados extraído de ensaios laboratoriais efetuados com 50 amostras de minério de ferro.

Foram realizados ensaios de separação densitária em líquido denso e ensaios de flotação, todos no Laboratório de Simulação e Controle de Processos Minerais da Universidade de São Paulo.

Alíquotas dos produtos foram encaminhadas à análise química. Os teores de ferro dos concentrados foram selecionados como os atributos na análise por agrupamento.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Definição do conceito

A terminologia *Cluster Analysis* (no Brasil, “Análise por Agrupamento” ou “Análise de Conglomerados”) foi utilizada pela primeira vez por Tryon em seu monograma intitulado *Cluster Analysis, 1939* (Hill, 2007), e engloba um conjunto de algoritmos e métodos utilizados para agrupar objetos similares, para um determinado conjunto de atributos pré-definidos. (Hair, 1998)

Bacher (1996) afirma que a ideia principal da análise de *cluster* é simples:

Encontrar *K clusters* (ou uma classificação que consista em *K clusters*) de modo que os objetos de um cluster sejam semelhantes entre si, enquanto objetos de diferentes clusters sejam dissimilares.

Everitt, (1981) enuncia sua definição relacionando as técnicas de agrupamento com o seguinte problema:

Dado um número de objetos ou indivíduos, cada qual descrito por um conjunto de medidas numéricas, elaborar um esquema de classificação para agrupar os objetos em um número de classes tais que os objetos pertencentes às classes sejam semelhantes em alguns aspectos e, distintos daqueles objetos que formam outras classes. O número de classes e as características de cada classe fazem parte do conjunto de determinações.

Já Gordon (1999) se expressa da seguinte forma:

A classificação se ocupa da investigação das relações dentro de um conjunto de "objetos", a fim de estabelecer se os dados podem ser validamente resumidos por um pequeno número de classes (ou clusters) de objetos similares.

Johann Bacher (2002) avalia as definições de Everitt e Gordon. Sobre Everitt, coloca que seu enunciado não compreende a aplicação das técnicas de agrupamento para variáveis, somente para casos. Bacher destaca ainda que o método pode buscar confirmações acerca de padrões previamente assumidos, enquanto Everitt sugere apenas o uso como análise exploratória. Sobre as colocações de Gordon, discute que a afirmação sobre o número de grupos ser pequeno é equivocada, pelo fato das técnicas por si só não pressuporem tal restrição. A avaliação posterior dos resultados é que fica tangível ao julgamento sobre a aceitação dos grupos obtidos ou à sua rejeição e reformulação da metodologia aplicada.

Jain (1988) alerta para a compreensão da diferença conceitual entre duas análises multivariadas distintas. A análise por agrupamento (não supervisionada) que não utiliza informações prévias sobre padrões ou rótulos, e a análise discriminante (supervisionada), que se ocupa em averiguar se um novo caso pertence a uma classificação pré-determinada, discriminando as variáveis que distinguem os grupos por meio de combinações lineares de variáveis independentes.

### 3.2 Aplicações

De acordo com Jain (1988) mesmo havendo um crescente interesse no uso de métodos de agrupamento em reconhecimento de padrões de imagem, processamento e recuperação de informação, o agrupamento tem uma história rica em outras disciplinas, tais como biologia, psiquiatria, psicologia, arqueologia, geologia, geografia e marketing. A importância e a natureza interdisciplinar do agrupamento são evidenciadas pela presença da vasta literatura. Hartigan (1975) fornece um resumo dos muitos estudos publicados até a época, relatando resultados de análise de *cluster*. Por exemplo, no campo da medicina, agrupar doenças, curas ou sintomas podem levar a taxonomias muito úteis. No campo da psiquiatria, o diagnóstico correto de conjuntos de sintomas tais como a esquizofrenia, paranoia, é essencial para a terapia de sucesso. Na arqueologia, os pesquisadores tentam estabelecer taxonomias de ferramentas de pedra e objetos funerários, aplicando técnicas de agrupamento.

Bacher (2002), afirma que até os anos 80 a discussão concentrava-se principalmente sobre as técnicas. No final dos anos 80, todo o processo de agrupamento, começando com a seleção de casos e variáveis e terminando com a validação dos *clusters*, tornou-se dominante.

As etapas de agrupamento propostas em seus trabalhos são as seguintes:

- 1) Seleção de casos, variáveis e métodos;
- 2) Aplicação dos métodos;
- 3) Avaliação dos resultados.

Esta última etapa inclui:

- a. Determinação do número de grupos, se desconhecida;
- b. Interpretação substancial de clusters;
- c. Testes de validação.

No âmbito do tratamento de minérios, a análise de agrupamento se relaciona não só com a simplificação controlada do conjunto de dados estudados, mas também pode ser utilizada, dentre inúmeras aplicações, para gerar mapas geometalúrgicos de similaridade de desempenho para modelos de blocos de lavra. Ou ainda, como ferramenta complementar em sistemas de controle automatizados de variáveis de processo. Imagens de formas de bolhas e cores de espumas de flotação podem ser utilizadas para comparar mapas de pixels, permitindo controle em tempo real das variáveis. (Cipriano *et al*, 1998). Classes podem ser definidas baseadas em variáveis petrofísicas de setores do maciço rochoso (por exemplo, amplitude e comprimento de ondas p, densidade, susceptibilidade magnética) estabelecendo, juntamente com modelos bem calibrados de regressão e análise de erros quadráticos, relações preditivas de características cominutivas, como britabilidade e moabilidade. (Vatandoost, 2010).

Antoniassi (2010) propõe o uso da técnica multivariada para análise de padrões difratométricos na caracterização de bauxitas. Mohamed (2005) utiliza o procedimento na definição de províncias geológicas de metais pesados presentes na costa do Mar Vermelho. Collyer (1973) investiga minas de tungstênio na América do Norte e estabelece, a partir da aplicação da técnica de agrupamento, áreas potenciais de exploração a partir de classificações de províncias metalogenéticas

### 3.3 Técnicas de agrupamento

Raramente é factível examinar todas as possibilidades de agrupamento de um conjunto de dados. Assim, uma grande variedade de algoritmos surgiu, almejando reduzir esforços computacionais e humanos na geração e ponderação dos resultados.

Qualquer que seja a técnica aplicada, definições são necessárias para se estabelecer uma rotina desejada, quais sejam:

- Métodos de associação entre clusters;
- Medidas de similaridade entre pares de pontos.

A similaridade entre dois clusters é expressa como uma função da distância entre dois pontos representativos destes clusters em um espaço  $p$ -dimensional. A determinação destes pontos (métodos associativos) e da medida da distância entre eles (medidas de similaridade) definem o método utilizado no agrupamento.

### 3.4 Métodos de Associação

Segundo Bacher (2007), os diversos métodos para a análise por agrupamento podem ser enquadrados em três grandes grupos:

- Análise por agrupamento hierárquico (aglomerativo ou divisivo);
- Análise por agrupamento não hierárquico;
- Análise por agrupamento probabilístico.

Técnicas hierárquicas podem ser aplicadas em tamanhos de amostra de 20 a 500 casos, dependendo da capacidade computacional disponível. Métodos não-hierárquicos, como o *K-Means* (*K-Médias*), requerem no mínimo amostras moderadas ( $n > 300$ ). Agrupamentos probabilísticos são significativos apenas para grandes amostras. ( $n > 3000$ )

Além das técnicas estatísticas de análise de cluster hierárquica e não-hierárquica, algoritmos evolutivos, agrupamentos *fuzzy*, redes neurais (mapas de Kohonen) podem ser empregadas para formação de agrupamentos. Modelos probabilísticos são discutidos por Fraley (2008).

O presente trabalho se ocupará somente do método de agrupamento hierárquico aglomerativo e do método particional *K-Means*.

### 3.4.1 Agrupamento Hierárquico Aglomerativo

Este método assume inicialmente que cada caso representa um cluster. Estes são progressivamente combinados, ou segundo Gordon (1999), amalgamados, utilizando matrizes de similaridade, recalculadas a cada iteração, até que todos os dados constituam um só grupo. Os resultados são geralmente visualizados através de um diagrama bidimensional chamado dendograma (Johnson, 1988), onde um eixo apresenta os rótulos de cada caso e o outro, as medidas de similaridade entre cada caso. A visualização das relações de proximidade e da estrutura de grupamento são percebidas facilmente na análise destes diagramas.

A análise de agrupamento hierárquico aglomerativo pode ser generalizada através do seguinte algoritmo:

- Considerar cada indivíduo como um *cluster*;
- Calcular a matriz de proximidade;
- Repetir:
  - Mesclar os dois *clusters* mais próximos;
  - Atualizar a matriz de proximidade utilizando uma medida de distância adequada;
  - Parar quando houver somente um *cluster*.

Existe uma variedade de métodos aglomerativos, que são caracterizados de acordo com o critério utilizado para definir as distâncias entre grupos. Entretanto, a maioria dos métodos parecem ser formulações alternativas de três grandes conceitos de agrupamento aglomerativo (Anderberg, 1973):

- 1) Métodos de ligação (*single-linkage*, *complete linkage*, *average-linkage*, *median-linkage*);
- 2) Métodos de centróide;
- 3) Métodos de soma de erros quadráticos ou variância (método de Ward).

Analogamente, Jain (1988) coloca que a maioria dos algoritmos de agrupamento hierárquico são derivações dos algoritmos *single-linkage* (Sneath e Sokal 1973), *complete-*

linkage (King, 1967), e mínima variância (Ward, 1963). A Figura 1 apresenta três métodos de agrupamento hierárquico (*single-linkage*, *complete-linkage*, *average-linkage*).

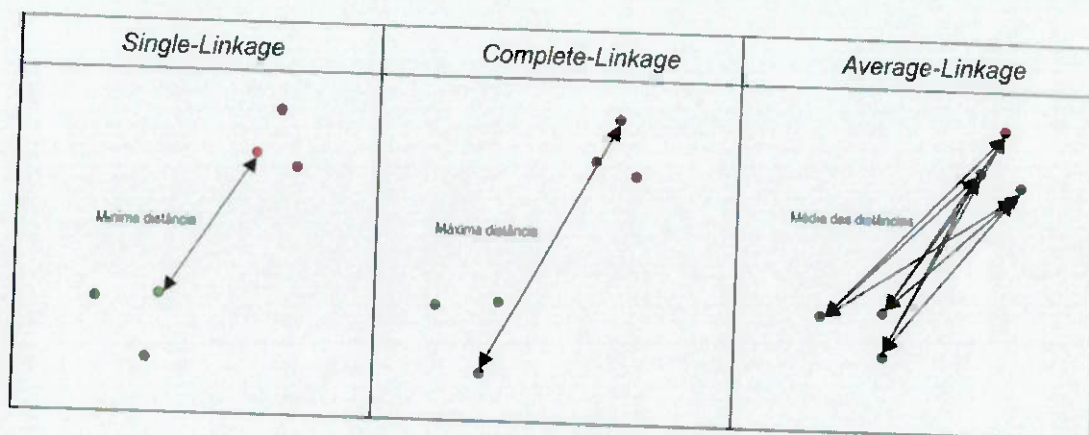


Figura 1 - Métodos de agrupamento hierárquico

Segundo Gordon (1999) diferentes resultados podem ser obtidos quando estratégias de agrupamento distintas são aplicadas a um mesmo conjunto de dados, sendo importante selecionar critérios apropriados aos dados selecionados.

No presente trabalho foram apresentados os métodos de ligação *single-linkage*, *complete-linkage* e *average-linkage*, e o método de Ward.

#### 3.4.1.1 Single-Linkage

O algoritmo *Single-Linkage*, ou Ligação ao Vizinho Mais Próximo (*Nearest Neighbour*), utiliza a menor distância entre os *clusters*.

Segundo Anderberg (1973) algumas características desse método são:

- Em geral, grupos muito próximos podem não ser identificados;
- Permite detectar grupos de formas não-elípticas;
- Apresenta pouca tolerância a ruído, pois tem tendência a incorporá-lo a um grupo já existente;

- Apresenta bons resultados tanto para distâncias Euclidianas quanto para outras distâncias;
- Tendência a formar longas cadeias (encadeamento).

### 3.4.1.2 Complete-Linkage

O algoritmo *Complete-Linkage* ou Ligação ao Vizinho Mais Distante (*Farthest Neighbour*). Ao contrário do algoritmo *Single-Linkage*, este determina a distância entre dois grupos de acordo com a maior distância entre um par de casos, sendo cada caso pertencente a um grupo distinto (Metz, 2006).

De acordo com Kaufman (1990) algumas características desse método são:

- Apresenta bons resultados tanto para distâncias Euclidianas quanto para outras distâncias;
- Tendência a formar grupos compactos;
- Os ruídos demoram a serem incorporados ao grupo.

### 3.4.1.3 Average-Linkage

O cálculo da dissimilaridade entre dois *clusters*  $R$  e  $Q$  é definido como a média de todas as dissimilaridades  $d(i,j)$ , onde  $i$  é qualquer objeto de  $R$ ,  $j$  é qualquer objeto de  $Q$  e  $|R|$  e  $|Q|$  denotam o número de elementos de cada *cluster*. A equação 1 a seguir apresenta a função de dissimilaridade utilizada no método *average-linkage*.

$$d(R, Q) = \frac{1}{|R||Q|} \sum_1^i \sum_1^j d(i, j) \quad (1)$$

De acordo com Kaufman (1990), este método apresenta as seguintes características:

- Menor sensibilidade a ruídos que os métodos de ligação por vizinho mais próximo e por vizinho mais distante;

- Apresenta bons resultados tanto para distâncias euclidianas quanto para outras distâncias;
- Tendência a formar grupos com número de elementos mais similares.

#### 3.4.1.4 Método de Ward

Este método utiliza uma análise de variância para avaliar as distâncias entre os conjuntos. A adesão é avaliada pelo cálculo da soma dos desvios quadrados a partir da média de um cluster. O critério para a fusão visa produzir o menor aumento possível na soma dos quadrados dos erros. Isto é, o método de Ward tenta minimizar a soma de quadrados dentro do grupo. Os grupos formados em cada passo são resultantes da solução com a menor soma de quadrados (Sharma, 1996). O método é baseado na noção de que os grupos devem ter, aproximadamente, forma elíptica. (Everitt, 2001)

Algumas características desse método são:

- Apresenta bons resultados tanto para distâncias euclidianas quanto para outras distâncias;
- Pode apresentar resultados insatisfatórios quando o número de elementos em cada grupo é praticamente igual;
- Tem tendência a combinar grupos com poucos elementos;
- Sensível à presença de *outliers*.

#### 3.4.2 Agrupamento Particional - *K-Means*

A análise por agrupamento particional *K-means*-(ou *K-médias*) é um processo iterativo onde é necessário determinar a priori o número de grupos desejados (*K*). Para cada grupo é calculado um valor para o centróide, utilizando a média dos parâmetros normalizados das amostras que compõem o grupo a cada iteração, como mostra a Figura 2.

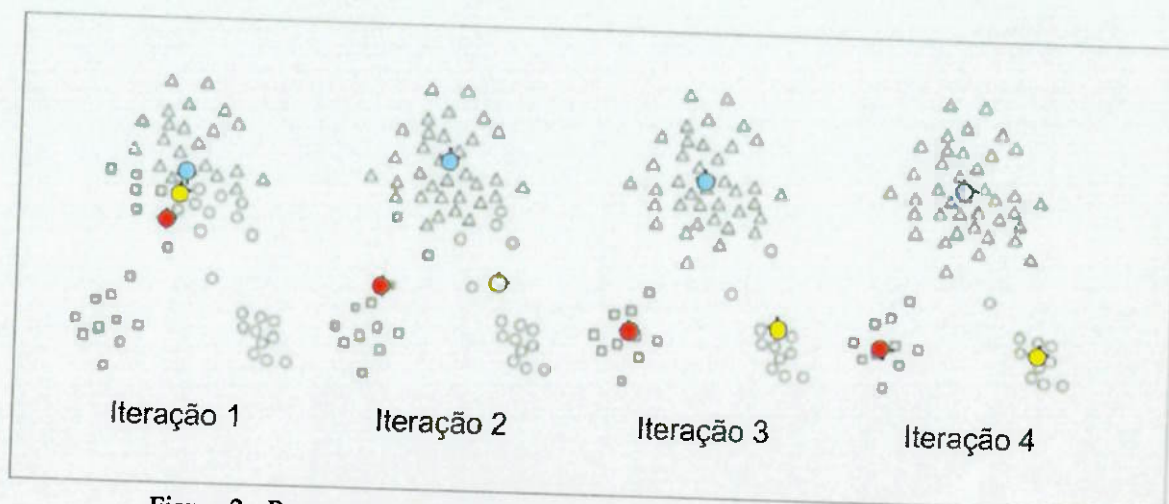


Figura 2 - Representação esquemática do método de agrupamento particional.

O agrupamento particional tem como algoritmo básico a seguinte sequência de operações:

- Arbitrar  $K$  pontos como centróides iniciais;
- Repetir:
  - Formar  $K$  *clusters* atribuindo todos os dados aos grupos cujos centróides estão mais próximos, medidas as proximidades (no presente caso) pela distância euclidiana;
  - Recalcular o centróide de cada *cluster*;
  - Parar quando os valores dos centróides não se alterarem após a  $i$ ésima iteração.

Para o agrupamento realizado, a medida de qualidade de cada grupo utilizada foi a soma dos erros quadráticos (SSE), descrita pela equação 2, onde,  $d(c_i, x)$  é a distância euclidiana entre o centróide ( $c_i$ ) e a amostra ( $x$ ).

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} d(c_i, x)^2 \quad (2)$$

Algumas características desse método são:

- Sensibilidade a ruídos, uma vez que um elemento com um valor extremamente alto pode distorcer a distribuição dos dados;
- Tendência a formar grupos esféricos;
- O número de grupos é o mesmo durante todo o processo;

- Inadequado para descobrir grupos com formas não convexas ou de tamanhos muito diferentes.

### 3.5 Medidas de similaridade

Como elucida Jonhson (2007), a aplicação das técnicas multivariadas de análise cluster pressupõem a definição de uma medida de proximidade ou, similaridade. Há uma parcela significativa de subjetividade envolvida na escolha de uma medida de similaridade. Considerações importantes incluem a natureza das variáveis (discretas, binárias, contínuas), as escalas de medição (nominal, ordinal, intervalar), e conhecimento prévio do assunto abordado.

Quando itens (casos) são agrupados, a proximidade é indicada por algum tipo de distância, enquanto variáveis são geralmente agrupadas com base em coeficientes de correlação ou medidas similares de associação.

Diversas formulações de medidas de distância são encontradas na literatura (Johnson, 2007), destacando-se:

- Distância Euclidiana;
- Distância de Manhattan;
- Distância de Chebychev;
- Métrica de Minkowski;
- Métrica de Canberra;
- Coeficiente de Czekanowski;

O presente trabalho utilizou a distância euclidiana ( $d(i,j)$ ), que pode ser expressa pela equação 3, onde:  $i = (xi_1, xi_2, \dots, xi_p)$  e  $j = (xj_1, xj_2, \dots, xj_p)$  são dois objetos p-dimensionais.

$$d(i,j) = \sqrt{|xi_1 - xj_1|^2 + |xi_2 - xj_2|^2 + \dots + |xi_p - xj_p|^2} \quad (3)$$

Alguns métodos impõem a utilização da distância euclidiana quadrática, que é obtida, como se intui, elevando-se a distância euclidiana ao quadrado.

Valores menores de similaridade indicam menores proximidades, enquanto valores menores de dissimilaridade representam maior similaridade. Na presente análise foram utilizadas relações de dissimilaridade, ou seja, quanto maiores os valores das distâncias euclidianas, maiores a dissimilaridades.

## 4 MATERIAIS E MÉTODOS

No presente trabalho foram estudadas 50 amostras de minério de ferro provenientes de testemunhos de sondagem, provenientes de um empreendimento mineiro em operação.

Cada uma das 50 amostras foi individualmente submetida à etapas sucessivas de britagem em britador de mandíbulas e peneirada na malha de 8 mm, de forma a que ao final toda a massa de cada amostra fosse passante na malha selecionada.

Ao final da etapa de britagem, cada amostra foi homogeneizada e quarteada para os ensaios subsequentes.

### 4.1 Análises Químicas

Para se determinar a composição química de cada amostra, uma alíquota resultante da homogeneização foi submetida a um novo estágio de britagem e peneiramento, até que todo o material passasse pela tela com abertura de 3,36 mm. O produto britado foi seco em estufa a 70° C até a secagem total do mesmo.

Além de Fe e SiO<sub>2</sub> foram determinados para cada uma das amostras, os teores de Al<sub>2</sub>O<sub>3</sub>, P, Mn, TiO<sub>2</sub>, CaO, K<sub>2</sub>O e perda ao fogo (PF). Contudo, somente os resultados de teores de ferro foram utilizados no presente trabalho.

### 4.2 Peneiramento

Uma alíquota de aproximadamente 6 kg resultante da homogeneização de cada amostra de minério de ferro foi encaminhada ao peneiramento à úmido, visando determinar a distribuição granulométrica dos produtos britados, bem como alíquotas de cada fração granulométrica para realização de ensaios e posterior desenvolvimento do teste padrão. Cada amostra foi,

portanto submetida a peneiramento a úmido, nas malhas de 1,0 mm (16# Tyler) e 0,15 mm (100# Tyler).

O peneiramento à úmido foi conduzido em bancada, utilizando-se três bacias preenchidas por água. Cada ensaio foi conduzido de modo a que em cada bacia, a peneira ficasse com a tela submersa e, através de movimentos circulares e ascendentes repetidos, o material fosse sucessivamente peneirado em cada uma das três bacias, até que a fração retida ficasse livre de partículas finas. Um indicador de eficiência elevada do processo foi manter a água da terceira bacia praticamente isenta de sólidos.

### **4.3 Separação em Líquido Denso**

Os ensaios de separação de minerais por densidade foram conduzidos sobre alíquotas das frações - 8,0 + 1,0 mm e - 1,0 + 0,15 mm.

As separações foram executadas com tetrabromoetano - TBE, cuja densidade nominal é  $2,96 \text{ g/cm}^3$ . No ensaio, as partículas mais densas que  $2,96 \text{ g/cm}^3$  afundam, enquanto que as menos densas que  $2,96 \text{ g/cm}^3$  flutuam. Assim, para cada amostra foram obtidos dois produtos: flutuado (leve) e afundado (pesado).

Por se tratar de frações relativamente grosseiras, os ensaios foram conduzidos em béqueres preenchidos por líquido denso. Alíquotas de aproximadamente 300 g foram adicionadas ao béquer contendo o líquido denso e então a polpa foi agitada por alguns segundos. Após a estabilização da separação, o material flutuado era removido do béquer com o auxílio de uma pequena peneira, enquanto que o material afundado foi encaminhado à filtragem. A Figura 3 mostra fotografias da separação em líquido denso.

Ao final de cada ensaio os produtos foram lavados com acetona, secos em estufa, pesados, quarteados e encaminhados para análises químicas.



Figura 3 - Fotografias dos ensaios de separação em líquido denso.

#### 4.4 Deslamagem

Após homogeneização e quarteamentos, uma alíquota com 1,5 kg da fração passante em 0,15 mm de cada uma das 50 amostras foi submetida à etapa de Deslamagem.

De acordo com o procedimento adotado, cada amostra foi disposta em um balde, juntamente com 8 litros de água, formando polpa com 15,8% de sólidos. A polpa foi então agitada por 5 minutos no próprio balde (1200 rpm), sendo então observado um período de sedimentação livre. O período de sedimentação livre foi determinado de acordo com o comportamento de cada amostra, e era iniciado assim que fosse visualizada a camada de sedimentação.

A polpa sobrenadante era retirada por sifonamento para um novo balde, com o auxílio de uma peneira de abertura 0,037 mm (400# Tyler).

O material sifonado foi encaminhado à secagem em estufa (100°C) e quarteamento, de forma a gerar alíquotas para análise química. A fração retida na peneira era então retornada ao balde, de maneira a controlar a recuperação em massa. O material sedimentado foi encaminhado à secagem em estufa a 70°C. Após secagem, cada amostra foi quarteada, com alíquotas encaminhadas à análises químicas, bem como à etapa de deslamagem secundária. A Figura 4 apresenta os procedimentos adotados nos ensaios de deslamagem.



Figura 4 - Fôtoğrafias dos ensaios de deslamagem.

Uma segunda etapa de deslamagem foi realizada mediante o mesmo procedimento. Após a secagem em estufa a 70°C, o material sedimentado de cada uma das amostras foi quarteado, com alíquotas encaminhadas a análises químicas, bem como à etapa de flotação.

#### 4.5 Flotação

Os ensaios de flotação foram realizados sobre as 50 amostras de *underflow* da deslamagem secundária, em estágios *rougher*, *scavenger*, *cleaner* e *recleaner*.. Foram empregadas alíquotas de 500 g de minério e 1,2 l de água para alimentação do estágio *rougher*, assim resultando em aproximadamente 30% de sólidos. O rejeito *rougher* foi integralmente encaminhado ao estágio *scavenger*. O concentrado *rougher* foi integralmente encaminhado ao estágio *cleaner*, cujo concentrado foi encaminhado a etapa *recleaner*. Todos os estágios foram conduzidos em cuba de 2,4 litros de capacidade. Ao final de cada ensaio, cinco produtos foram obtidos: rejeito *cleaner*, concentrado e rejeito *scavenger*, concentrado e rejeito *recleaner*.

O fluxograma dos ensaios de flotação segue apresentado na Figura 5.

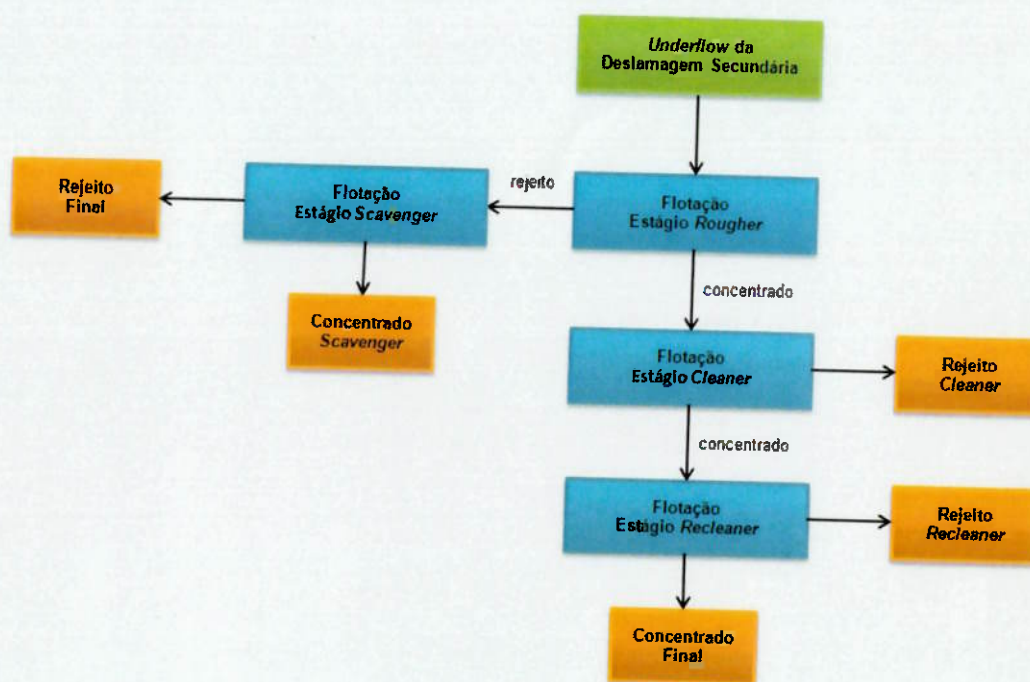


Figura 5 - Fluxograma dos estágios dos ensaios de flotação.

No agrupamento foram considerados os valores obtidos para o concentrado *recleaner*, também identificado como concentrado final.

Uma máquina da marca Denver, apresentada na Figura 6, foi empregada para condução de todos os ensaios de flotação.



Figura 6 - Máquina marca Denver empregada nos ensaios de flotação.

A rotação da máquina foi ajustada em 1200 rpm para o condicionamento e a flotação. A porcentagem de sólidos no condicionamento foi ajustada de maneira que o anel da máquina fosse coberto por polpa, garantindo a boa agitação da polpa durante esta etapa. Após o condicionamento a água era adicionada até preencher o volume adequado para flotação.

Em todos os ensaios de flotação foram utilizados o coletor amina e, como depressor, amido. Os períodos de condicionamento da polpa com o depressor (amido) e coletor (amina) foram, respectivamente de cinco e três minutos. Não foi necessário o uso de espumantes, uma vez que a amina fornece a espuma necessária. A adição dos reagentes foi volumétrica, com base em dosagens (g/t) de alimentação do ensaio.

Os procedimentos estabelecidos para preparação dos reagentes empregado nos ensaios de flotação foram padronizados para todas as etapas de flotação.

Para a preparação do depressor, o procedimento consistiu em adicionar 5 g do reagente a um béquer junto, bem como 50 ml de água. O amido foi gelatinizado com solução de cloreto de sódio a 10%. A proporção entre amido e cloreto de sódio foi de 5:1, respectivamente. Posteriormente a solução era diluída a 500 ml, obtendo-se portanto uma solução com concentração de 1%.

Procedimento análogo foi empregado para preparação de solução de coletor, neste caso para massa de 2 g do coletor foram adicionados 0,22 de óleo diesel (7,5:1) e o volume de 100 ml foi preenchido por água, obtendo-se portanto, uma solução com concentração de 2% de amina. A Tabela 1 sumariza as adições de reagentes para cada fluxo do ensaio de flotação.

Tabela 1 - Condições estabelecidas para os ensaios de flotação.

| Reagente | Preparação                              | Dosagem (g/t)  |                  |                |                  |
|----------|-----------------------------------------|----------------|------------------|----------------|------------------|
|          |                                         | <i>rougher</i> | <i>scavenger</i> | <i>cleaner</i> | <i>recleaner</i> |
| Amido    | 5 : 1<br>(amido/soda)<br>diluído a 1%   | 400            | 300              | 0              | 0                |
| Amina    | 7,5 : 1<br>(amina/óleo)<br>diluído a 2% | 150            | 0                | 50             | 50               |

O pH da flotação foi monitorado e mantido em 10,5 mediante adição de solução preparada à 1% (m/v) de hidróxido de sódio. A flotação foi conduzida até a exaustão da espuma.

#### 4.6 Agrupamento de Amostras

O agrupamento de amostras visou determinar semelhanças entre as amostras individuais, com base em características tecnológicas. O recurso selecionado para a classificação das amostras de minério ferro foi a análise por agrupamento (*cluster analysis*), a qual permitiu associar diferentes amostras em grupos de semelhança. Foi utilizado o *software* IBM-SPSS v20 para realização dos cálculos.

As variáveis utilizadas como atributos de cada amostra foram:

- 1) Teor de ferro no concentrado denso da fração - 8,00 + 1,00 mm.
- 2) Teor de ferro no concentrado denso da fração - 1,00 + 0,15 mm.
- 3) Teor de ferro no concentrado final da flotação.

Os parâmetros escolhidos tiveram por objetivo identificar grupos que apresentassem desempenhos distintos com relação aos três teores de ferro presentes nos concentrados laboratoriais.

A técnica hierárquica aglomerativa foi utilizada preliminarmente para geração do dendograma e visualização geral das proximidades entre amostras, com o objetivo de definir o número de grupos a serem formados, bem como identificar possíveis *outliers*.

A seleção do método considerou o número de amostras em cada grupo gerado nas análises preliminares, bem como a diferença estatística para as variáveis avaliadas.

A técnica particional (*K-Means*) foi utilizada para refinamento da solução.

Para ambos os métodos utilizou-se a padronização *z-scores* e como medida de proximidade a distância euclidiana quadrática.

Foram realizados cálculos de medidas centrais (média e mediana) e medidas de dispersão (variância, desvio-padrão, mínimo, máximo, amplitude, percentis e distância interquartilica). Coeficientes de assimetria e de curtose com a correção de Fisher foram utilizados como subsídio para a análise da forma da distribuição.

Na apresentação dos dados foram utilizados: *boxplots* para visualização de medidas centrais (mediana), medidas de dispersão (quartis) e *outliers*.

A validação dos cálculos foi feita através de um teste de hipóteses para comparação entre  $K$  médias, utilizando-se o teste  $t$  para comparação entre médias, acompanhado do teste de Levene para determinação da homogeneidade de variâncias e da correção de graus de liberdade de Aspin-Welch.

## 5 RESULTADOS E DISCUSSÕES

As variáveis utilizadas como atributos das 50 amostras foram:

- 1) Teor de ferro do concentrado da separação em líquido denso (-8 + 1, mm), chamada indistintamente neste documento de C1.
- 2) Teor de ferro do concentrado da separação em líquido denso (-1 + 0,150 mm), chamada indistintamente neste documento de C2.
- 3) Teor de ferro do concentrado de flotação, chamada indistintamente neste documento de C3.

Os teores de ferro dos concentrados de meio denso e de flotação encontram-se apresentados na Tabela 2, a seguir.

Tabela 2 - Teores de ferro nos concentrados.

| Amostra | Teores de Fe nos Concentrados (%) |                               |          |
|---------|-----------------------------------|-------------------------------|----------|
|         | Líquido denso (- 8 + 1 mm)        | Líquido denso (- 1 + 0,15 mm) | Flotação |
|         | C1                                | C2                            | C3       |
| 1       | 49,1                              | 61,1                          | 58,8     |
| 2       | 58,3                              | 57,5                          | 52,4     |
| 3       | 65,8                              | 65,6                          | 55,6     |
| 4       | 63,6                              | 62,3                          | 58,9     |
| 5       | 53,8                              | 62,0                          | 51,5     |
| 6       | 57,5                              | 62,0                          | 58,7     |
| 7       | 63,9                              | 64,8                          | 50,9     |
| 8       | 63,6                              | 63,5                          | 51,1     |
| 9       | 65,8                              | 64,9                          | 54,6     |
| 10      | 59,2                              | 57,0                          | 56,0     |
| 11      | 55,6                              | 53,8                          | 56,1     |
| 12      | 62,4                              | 65,9                          | 50,9     |
| 13      | 61,5                              | 65,3                          | 58,5     |
| 14      | 62,9                              | 63,6                          | 59,7     |
| 15      | 59,6                              | 60,0                          | 57,7     |
| 16      | 64,3                              | 64,3                          | 42,6     |
| 17      | 52,7                              | 60,2                          | 58,4     |
| 18      | 56,2                              | 57,3                          | 50,8     |
| 19      | 61,1                              | 64,1                          | 50,0     |
| 20      | 57,2                              | 58,9                          | 59,7     |
| 21      | 59,0                              | 56,0                          | 55,5     |
| 22      | 52,1                              | 63,0                          | 52,4     |
| 23      | 67,8                              | 67,7                          | 68,7     |
| 24      | 65,4                              | 63,8                          | 61,5     |
| 25      | 63,6                              | 65,5                          | 62,1     |
| 26      | 66,0                              | 66,0                          | 58,8     |
| 27      | 59,6                              | 62,8                          | 59,4     |
| 28      | 59,1                              | 59,6                          | 49,2     |
| 29      | 56,7                              | 61,9                          | 59,4     |
| 30      | 66,1                              | 64,5                          | 64,6     |
| 31      | 62,7                              | 65,1                          | 68,4     |
| 32      | 64,7                              | 64,0                          | 60,8     |
| 33      | 63,9                              | 62,3                          | 61,1     |
| 34      | 62,5                              | 63,0                          | 66,5     |
| 35      | 57,7                              | 57,5                          | 55,2     |
| 36      | 60,6                              | 65,4                          | 52,9     |
| 37      | 51,3                              | 51,6                          | 62,7     |
| 38      | 68,2                              | 67,6                          | 56,9     |
| 39      | 55,1                              | 64,5                          | 53,6     |
| 40      | 64,1                              | 66,4                          | 57,9     |
| 41      | 60,3                              | 61,6                          | 49,6     |
| 42      | 66,5                              | 62,7                          | 55,4     |
| 43      | 57,2                              | 54,6                          | 44,8     |
| 44      | 67,8                              | 67,9                          | 61,6     |
| 45      | 64,3                              | 64,2                          | 63,7     |
| 46      | 60,8                              | 62,4                          | 63,5     |
| 47      | 65,8                              | 64,9                          | 50,9     |
| 48      | 68,5                              | 67,9                          | 49,2     |
| 49      | 58,1                              | 58,1                          | 46,0     |
| 50      | 65,9                              | 66,3                          | 61,6     |

Foram realizados cálculos de medidas centrais (média e mediana) e medidas de dispersão (variância, desvio-padrão, mínimo, máximo, amplitude, percentis e distância interquartilica). Coeficientes de assimetria e de curtose com a correção de Fisher foram utilizados como subsídio para a análise da forma da distribuição. Os resultados para as variáveis C1, C2 e C3 estão apresentados na Tabela 3, a seguir.

Tabela 3 - Análise descritiva das variáveis C1, C2 e C3.

| Estatística                                |                 | Teor de Fe (%) |        |        |
|--------------------------------------------|-----------------|----------------|--------|--------|
|                                            |                 | C1             | C2     | C3     |
| Média                                      |                 | 61,11          | 62,46  | 56,54  |
| Intervalo de 95% de confiança para a média | Limite inferior | 59,76          | 61,38  | 54,86  |
|                                            | Limite superior | 62,46          | 63,54  | 58,21  |
| Média - distribuição truncada (5%)         |                 | 61,30          | 62,68  | 56,58  |
| Mediana                                    |                 | 61,95          | 63,25  | 57,30  |
| Variância                                  |                 | 22,68          | 14,47  | 34,58  |
| Desvio-padrão amostral                     |                 | 4,76           | 3,80   | 5,88   |
| Mínimo                                     |                 | 49,10          | 51,60  | 42,60  |
| Máximo                                     |                 | 68,50          | 67,90  | 68,70  |
| Amplitude                                  |                 | 19,40          | 16,3   | 26,10  |
| Distância Interquartilica                  |                 | 7,22           | 5,00   | 9,48   |
| Coeficiente de assimetria de Fisher        |                 | -0,546         | -0,924 | -0,137 |
| Coeficiente de curtose de Fisher           |                 | -0,355         | 0,447  | -0,264 |

As distribuições das variáveis C1, C2 e C3 estão apresentadas em histogramas e diagramas de normalidade  $Q-Q$ , no Apêndice A.

Foram realizados testes de normalidade que buscaram atestar se as variáveis C1, C2 e C3 possuem distribuição normal. Como o número de amostras é igual a 50, utilizou-se o teste de Shapiro-Wilk. Valores de significância maiores que 0,05 sugeriram a aceitação da hipótese nula de que a distribuição é normal. A Tabela 4 apresenta os valores encontrados.

Tabela 4 - Testes de normalidade.

| Variáveis                                     | Teste de Shapiro-Wilk |                    |               |
|-----------------------------------------------|-----------------------|--------------------|---------------|
|                                               | Estatística           | Graus de liberdade | Significância |
| Teor de Fe no Concentrado - 8 + 1 mm (C1)     | 0,961                 | 50                 | 0,094         |
| Teor de Fe no Concentrado - 1 + 0,150 mm (C2) | 0,930                 | 50                 | 0,006         |
| Teor de Fe no Concentrado Flotação (C3)       | 0,986                 | 50                 | 0,820         |

O teste de normalidade de Shapiro-Wilk é robusto para o número de amostras ( $N = 50$ ) e atestou a normalidade das distribuições de C1 (Shapiro-Wilk 0,961;  $df = 50$ ;  $p = 0,094$ ) e de C3 (Shapiro-Wilk 0,986;  $df = 50$ ;  $p = 0,820$ ) e a não-normalidade de C2 (Shapiro-Wilk 0,930;  $df = 50$ ;  $p = 0,006$ ). As reduções  $df$  e  $p$  utilizadas, se referiram às expressões *degrees of freedom* (graus de liberdade) e *p-value* (equivalente a “significância”).

Como para pelo menos uma das variáveis (C2) não se pôde aceitar a hipótese nula de normalidade, o estudo de correlação realizado utilizou o coeficiente de Spearman (*Spearman's rho*), mais robusto que o coeficiente de correlação de Pearson quando aplicado à distribuições não paramétricas, e os resultados estão apresentados na Tabela 5 a seguir.

Tabela 5 - Matriz de correlação de Spearman.

| Matriz de correlação de Spearman |                 |       |       |       |
|----------------------------------|-----------------|-------|-------|-------|
|                                  |                 | C1    | C2    | C3    |
| C1                               | Spearman's rho  | 1,000 | 0,764 | 0,214 |
|                                  | Sig. (bicaudal) | -     | 0,000 | 0,136 |
|                                  | N               | 50    | 50    | 50    |
| C2                               | Spearman's rho  | 0,764 | 1,000 | 0,187 |
|                                  | Sig. (bicaudal) | 0,000 | -     | 0,192 |
|                                  | N               | 50    | 50    | 50    |
| C3                               | Spearman's rho  | 0,214 | 0,187 | 1,000 |
|                                  | Sig. (bicaudal) | 0,136 | 0,192 | -     |
|                                  | N               | 50    | 50    | 50    |

As variáveis C1 e C2 apresentaram correlação moderada (Spearman's rho = 0,764, sig. < 0,001), no entanto, não suficiente para exclusão de uma das variáveis no procedimento de agrupamento.

Assim, as 50 amostras foram agrupadas segundo os três parâmetros pré-definidos. O dendograma produzido pelo agrupamento hierárquico aglomerativo, utilizando-se a distância euclidiana quadrática e o método de Ward está apresentado no Apêndice A.

A partir do dendograma, constatou-se a existência de dois grandes grupos de amostras. Foi então realizado o refinamento da solução através da técnica particional K-Means, impondo ao algoritmo dois centróides iniciais ( $K=2$ ). Os grupos formados e as

distâncias de seus elementos aos centróides calculados estão apresentados na Tabela 6, a seguir.

Tabela 6 - Composição dos grupos e distâncias dos elementos aos centróides.

| Amostra | Grupo | Distância ao centróide do grupo | Amostra | Grupo | Distância ao centróide do grupo |
|---------|-------|---------------------------------|---------|-------|---------------------------------|
| 1       | 1     | 9,5                             | 3       | 2     | 3,5                             |
| 2       | 1     | 2,7                             | 4       | 2     | 2,7                             |
| 5       | 1     | 4,6                             | 7       | 2     | 7,9                             |
| 6       | 1     | 5,7                             | 8       | 2     | 7,8                             |
| 10      | 1     | 4,1                             | 9       | 2     | 4,4                             |
| 11      | 1     | 6,2                             | 12      | 2     | 8,2                             |
| 15      | 1     | 4,9                             | 13      | 2     | 3,0                             |
| 16      | 1     | 14,2                            | 14      | 2     | 2,2                             |
| 17      | 1     | 6,3                             | 23      | 2     | 10,9                            |
| 18      | 1     | 3,6                             | 24      | 2     | 3,1                             |
| 19      | 1     | 7,3                             | 25      | 2     | 3,5                             |
| 20      | 1     | 6,0                             | 26      | 2     | 1,9                             |
| 21      | 1     | 4,4                             | 27      | 2     | 5,3                             |
| 22      | 1     | 6,1                             | 30      | 2     | 6,1                             |
| 28      | 1     | 5,0                             | 31      | 2     | 9,8                             |
| 29      | 1     | 6,2                             | 32      | 2     | 2,2                             |
| 35      | 1     | 2,6                             | 33      | 2     | 3,5                             |
| 37      | 1     | 13,2                            | 34      | 2     | 8,2                             |
| 39      | 1     | 5,4                             | 36      | 2     | 7,0                             |
| 41      | 1     | 5,8                             | 38      | 2     | 5,0                             |
| 43      | 1     | 10,1                            | 40      | 2     | 1,8                             |
| 49      | 1     | 7,9                             | 42      | 2     | 4,5                             |
|         |       |                                 | 44      | 2     | 5,3                             |
|         |       |                                 | 45      | 2     | 5,0                             |
|         |       |                                 | 46      | 2     | 6,5                             |
|         |       |                                 | 47      | 2     | 8,0                             |
|         |       |                                 | 48      | 2     | 10,8                            |
|         |       |                                 | 50      | 2     | 3,5                             |

Assim, vinte de duas amostras (22) foram alocadas no primeiro grupo e vinte e oito (28) alocadas no segundo grupo. Os resultados descritivos estão apresentados na Tabela 7 a seguir.

Tabela 7 - Análise descritiva por grupo das variáveis C1, C2 e C3.

| Estatística                                   | Teor de Fe (%)       |        |        |                      |        |        |
|-----------------------------------------------|----------------------|--------|--------|----------------------|--------|--------|
|                                               | Grupo 1: 22 amostras |        |        | Grupo 2: 28 amostras |        |        |
|                                               | C1                   | C2     | C3     | C1                   | C2     | C3     |
| Média                                         | 56,87                | 59,39  | 53,69  | 64,44                | 64,87  | 58,78  |
| Intervalo de 95% de<br>confiança para a média | Limite inferior      | 55,32  | 57,83  | 51,35                | 63,53  | 64,20  |
|                                               |                      | 58,43  | 60,95  | 56,02                | 65,35  | 65,53  |
| Média - distribuição truncada (5%)            |                      | 56,90  | 59,53  | 53,81                | 64,47  | 64,84  |
| Mediana                                       |                      | 57,35  | 59,80  | 54,40                | 64,20  | 64,90  |
| Variância                                     |                      | 12,30  | 12,37  | 27,82                | 5,46   | 2,95   |
| Desvio-padrão amostral                        |                      | 3,51   | 3,52   | 5,27                 | 2,34   | 1,72   |
| Mínimo                                        |                      | 49,10  | 51,60  | 42,60                | 59,60  | 62,30  |
| Máximo                                        |                      | 64,30  | 64,50  | 62,70                | 68,50  | 67,90  |
| Amplitude                                     |                      | 15,20  | 12,9   | 20,10                | 8,90   | 5,6    |
| Distância Interquartilica                     |                      | 4,35   | 4,78   | 8,58                 | 3,22   | 2,45   |
| Coefficiente de assimetria de Fisher          |                      | -0,335 | -0,439 | -0,416               | -0,142 | 0,217  |
| Coefficiente de curtose de Fisher             |                      | 0,364  | -0,377 | -0,457               | -0,516 | -0,784 |
|                                               |                      |        |        |                      |        | -0,691 |

Os resultados para as variáveis C1, C2 e C3, foram sumarizados através de diagramas *boxplot*, apresentados no Apêndice A.

Foram realizados testes de normalidade que buscaram atestar se as variáveis C1, C2 e C3 possuem distribuição normal considerando separadamente os dois grupos. Como o número de amostras é igual inferior a 50, utilizou-se o teste de Shapiro-Wilk. Valores com significância maior que 0,05 sugeriram a aceitação da hipótese nula de que determinada distribuição é normal. A Tabela 8 apresenta os valores encontrados.

Tabela 8 - Testes de normalidade.

| Variáveis                                     | Grupo | Teste de Shapiro-Wilk |                    |               |
|-----------------------------------------------|-------|-----------------------|--------------------|---------------|
|                                               |       | Estatística           | Graus de liberdade | Significância |
| Teor de Fe no Concentrado - 8 + 1 mm (C1)     | 1     | 0,974                 | 22                 | 0,805         |
|                                               | 2     | 0,975                 | 28                 | 0,711         |
| Teor de Fe no Concentrado - 1 + 0,150 mm (C2) | 1     | 0,965                 | 22                 | 0,592         |
|                                               | 2     | 0,951                 | 28                 | 0,216         |
| Teor de Fe no Concentrado Flotação (C3)       | 1     | 0,968                 | 22                 | 0,672         |
|                                               | 2     | 0,964                 | 28                 | 0,440         |

Todas as variáveis apresentaram distribuições normalmente distribuídas, considerando-se separadamente os grupos 1 e 2.

A validação dos dados foi, então, feita através de um teste de comparação de duas médias com dados não emparelhados, para grupos com números de amostras distintos. Como no presente caso foram encontrados somente dois grupos, testes com a estatística *t* poderiam ser utilizados em detrimento do teste de análise de variâncias (*ANOVA*). No entanto, ambos os testes pressupõem: a ausência de heterocedasticidade (diferenças entre variâncias), que deve ser verificada pelo teste de Levene; e semelhança entre o número de amostras de cada grupo, que deve ser tratada pela correção de Aspin-Welch. (Almeida, 2006). O teste *t* e os parâmetros utilizados nos cálculos estão apresentados na Tabela 9, a seguir.

Tabela 9 - Comparação entre grupos.

| Variáveis | Variâncias | Teste de Levene |       | Teste <i>t</i> para comparação entre médias |                  |                 |
|-----------|------------|-----------------|-------|---------------------------------------------|------------------|-----------------|
|           |            | F               | Sig.  | <i>z</i>                                    | df (Aspin-Welch) | Sig. (bicaudal) |
| C1        | Iguais     | 2,291           | 0,137 | -9,134                                      | 48               | 0,000           |
|           | Distintas  |                 |       | -8,713                                      | 34,9             | 0,000           |
| C2        | Iguais     | 12,586          | 0,001 | -7,228                                      | 48               | 0,000           |
|           | Distintas  |                 |       | -6,702                                      | 28,8             | 0,000           |
| C3        | Iguais     | 0,000           | 0,990 | -3,337                                      | 48               | 0,002           |
|           | Distintas  |                 |       | -3,348                                      | 45,8             | 0,002           |

O teste de Levene tem como hipótese nula a igualdade de variâncias entre os grupos. Se a significância do teste for menor que 0,05 será difícil atestar que a diferença entre as variâncias se deve ao acaso e deve-se rejeitar a hipótese nula. Se a significância do teste for maior que 0,05, não se pode rejeitar a hipótese nula e pode-se supor a igualdade de variâncias, o que aumenta o poder dos testes de comparação entre médias.

O teste de Levene revelou heterocedasticidade para a variável C2 e homocedasticidade para as variáveis C1 e C3. Os dados utilizados estão grifados em negrito na Tabela 9.

O teste *t* revelou que os grupos (1 e 2) diferem para todas as três variáveis (C1, C2 e C3) a um nível de significância de 1%.

Assim é possível concluir que o grupo 1, formado por 22 amostras, possuiu menores teores de ferro (menor qualidade) para as variáveis C1, C2 e C3, e o grupo 2, constituído por 28 amostras, possuiu maiores teores de ferro (maior qualidade) para as variáveis C1, C2 e C3, validando o procedimento de agrupamento adotado.

## 6 CONCLUSÕES

Foram apresentadas características e aplicações de técnicas de análise por agrupamento, incluindo a utilização na área de tratamento mineral.

A análise por agrupamento foi realizada a partir dos teores de concentrados de minério de ferro. Ensaios padronizados de separação em líquido denso e de flotação foram realizados em escala de bancada para 50 amostras. A análise por agrupamento resultou na identificação de dois grupos entre

O primeiro (grupo 1), com 22 amostras, apresentou teores de ferro médios inferiores para as três variáveis, quais sejam: teor de ferro no concentrado denso da fração - 8,00 + 1,00 mm (56,9%); teor de ferro no concentrado denso da fração - 1,00 + 0,15 mm (59,4%); e teor de ferro no concentrado final da flotação (53,7%). Este grupo pode ser rotulado como o de qualidade inferior.

O segundo (grupo 2), com 28 amostras, apresentou teores de ferro médios superiores para as três variáveis, quais sejam: teor de ferro no concentrado denso da fração - 8,00 + 1,00 mm (64,4%); teor de ferro no concentrado denso da fração - 1,00 + 0,15 mm (64,9%); e teor de ferro no concentrado final da flotação (58,8%). Este grupo pode ser rotulado como o de qualidade superior.

Tais grupos podem ser fisicamente consumados através de homogeneizações, e ensaios mais detalhados podem ser efetuados sobre apenas dois grupos.

A aplicação da técnica descrita resultou, portanto, em dois grupos distintos de amostras, conforme apontam os resultados médios. A técnica apresentada é, portanto, útil em programas de geometurgia e na definição de grupos de comportamentos distintos com relação a ensaios de caracterização tecnológica ou desempenho industrial.

## 7 REFERÊNCIAS

- ALMEIDA, A.; ELIAN, S.; NOBRE, J. **Modificações e alternativas aos testes de Levene e de Brown & Forsythe para igualdade de variâncias e médias.** Revista Colombiana de Estadística. v.31, nº 2, pag. 241 a 260. Bogotá, 2008
- ANDERBERG, Michael R. **Cluster analysis for applications.** New York: Academic Press, 1973.
- ANTONIASSI, J.L. **A difração de raios-X com o método Rietveld aplicada a bauxitas do Porto de Trombetas, PA.** 102 p. Dissertação de mestrado – Escola Politécnica da Universidade de São Paulo, São Paulo, 2010.
- CHAVES, A.P.; CORRÊA, F.D. **Dimensionamento de pequenos e médios projetos de mineração.** Brasil Mineral - nº 56, 1988, p. 26-34.
- FRALEY, C.; RAFTERY A. E. **How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis.** Department of Statistics, University of Washington, USA, 2008.
- CIPRIANO, A. et al. **A real time visual sensor for supervision of flotation cells.** Minerals Engineering, 11(6), pag. 489-499.
- COLLYER, P. L.; MERRIAM, D. F. **An application of cluster analysis in mineral exploration.** Journal of the International Association for Mathematical Geology. Volume 5, Issue 3, 1973, pag 213-223.
- EVERITT, B. **Cluster analysis.** Segunda edição. Londres-Nova Iorque, 1981.
- EVERITT, B.; LANDAU, S.; LESSE, M. **Cluster analysis.** Quarta edição. London-New York, 2001.
- GORDON, A. D. **Classification.** Chapman and Hall Ed., 1981.

- GORDON, A. D. **Classification**. 2nd edition. London and others, 1999.
- HAIR, J. et al. **Multivariate Data Analysis**. New Jersey, Prentice Hall, 1998.
- HARTIGAN, J. A.: **Clustering Algorithms (Probability & Mathematical Statistics)**. Wiley, New York, 1975.
- HILL, T.; LEWICKI, P. **Statistics methods and applications**, StatSoft, Tulsa, OK, 2007.
- JAIN, A. K.; DUBES, R.C. **Algorithms for Clustering Data**. Englewood Cliffs, New Jersey, 1988.
- JOHNSON, R. A. and WICHERN, D. W. **Applied Multivariate Analysis**. 2nd ed. Englewood Cliffs, NJ, Prentice-Hall, 1988.
- KAUFMAN, L.; ROUSSEEUW, P. J. **Finding groups in data: an introduction to cluster analysis**. Wiley, New York, 1990.
- KING, B. **Step-wise clustering procedures**. Journal of the American Statistical Association, 1967. v. 69, pag 86-101.
- METZ, J. **Interpretação de Clusters Gerados por Algoritmos de *Clustering* Hierárquico**. Dissertação de Mestrado, USP, São Carlos - SP, 2006.
- MOHAMED. A, **Cluster Analisis and Mineral Provenances of Recent Sediments and Their Relation to the Continental Margin Activity Along the Red Sea Coast**. Egypty. Egyptyan Journal of Aquatic Research. 1687-4285; vol.31 n° 2, 2005.
- SHARMA, S. **Applied multivariate techniques**. John Wiley & Sons, New York, 1996.
- SNEATH, P. H.; SOKAL, R. R. **Numerical taxonomy: The principles and practice of numerical classification**. San Francisco: W.H. Freeman, 1973. 573p.

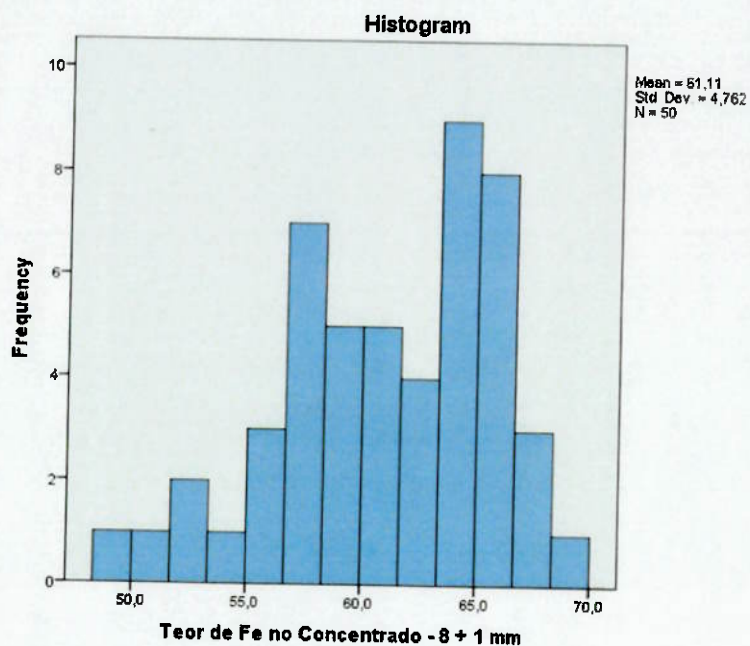
VATANDOOST, A. **Petrophysical characterization of comminution behavior..** PhD thesis, University of Tasmania, 2010.

WARD J. H. **Hierarchical Grouping to Optimize an Objective Function.** American Statistical Association Journal, March 1963.

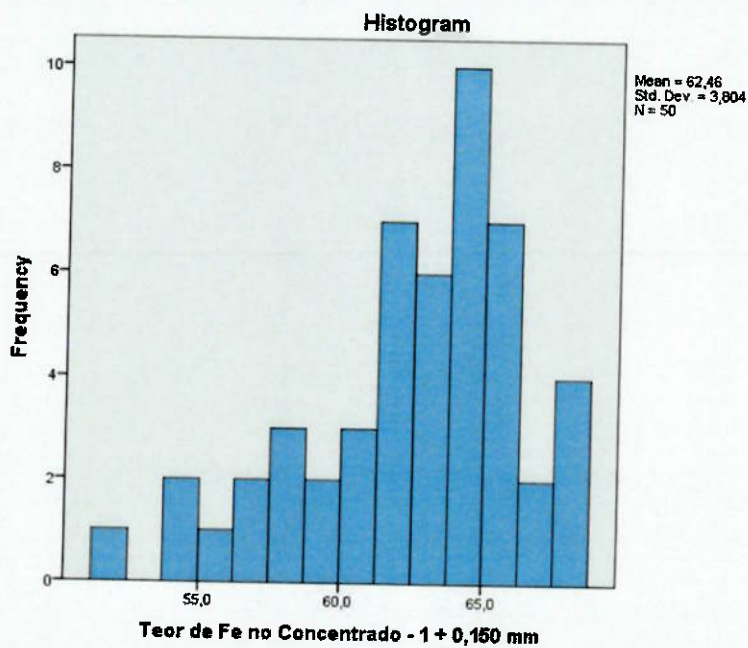
## APÊNDICE A – Gráficos

### Histogramas de frequência – Teores de ferro de 50 amostras

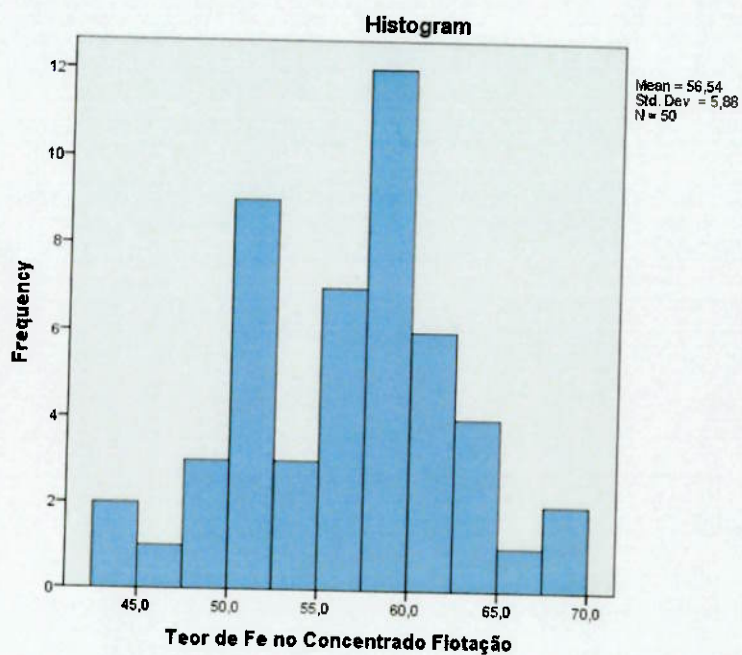
Concentrado da separação em líquido denso da fração - 8 + 1 mm



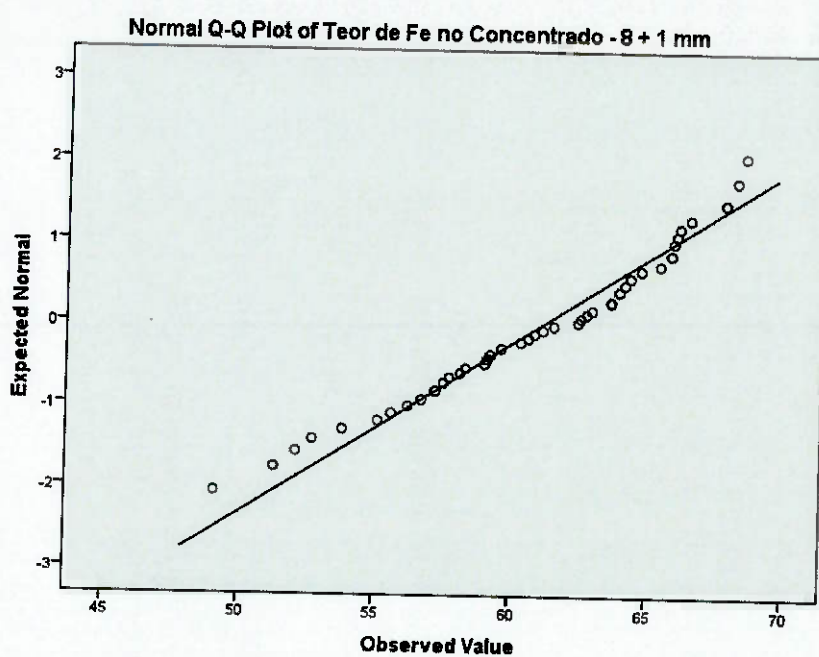
Concentrado da separação em líquido denso da fração - 1 + 0,150 mm



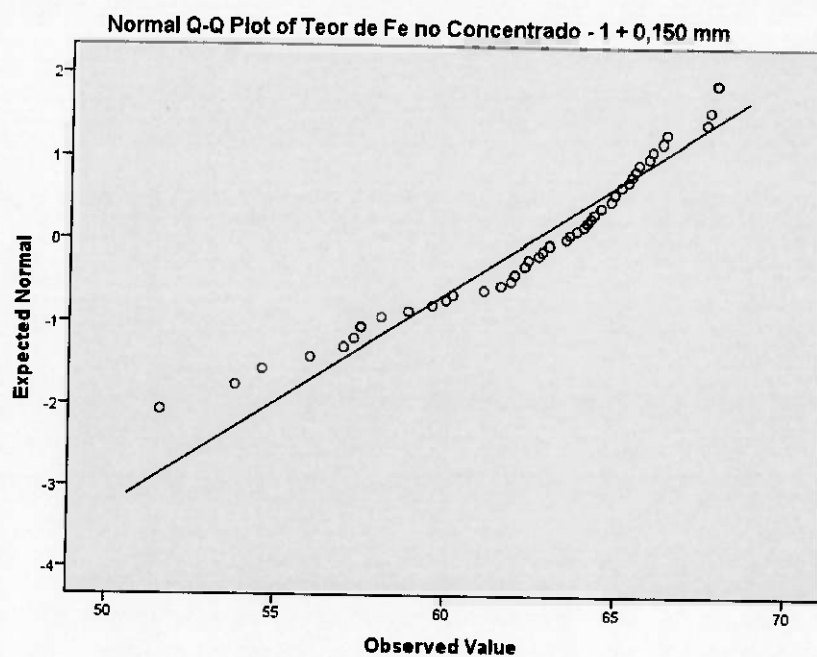
## Concentrado da flotação.

Gráficos  $Q-Q$  – Teores de ferro de 50 amostras

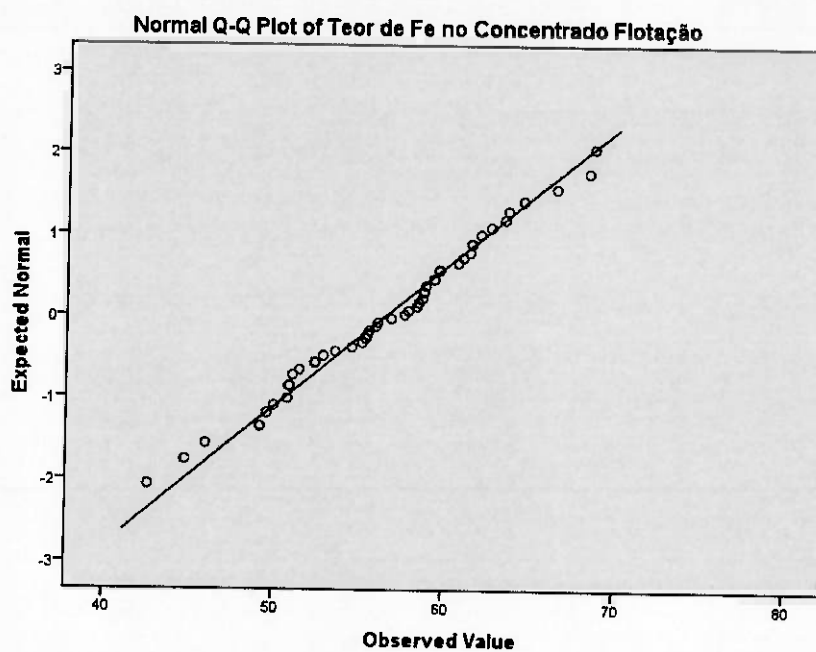
Concentrado da separação em líquido denso da fração - 8 + 1 mm



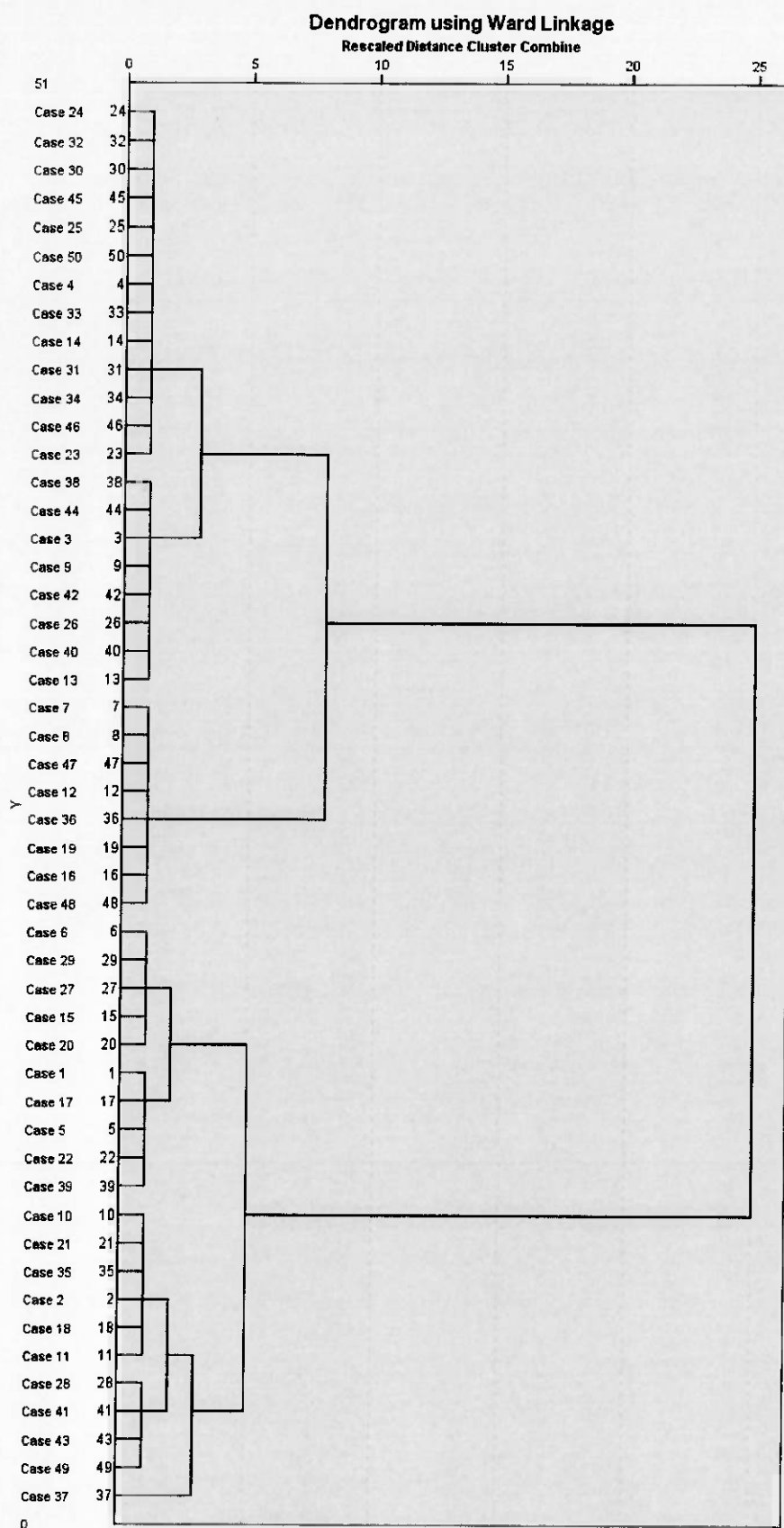
Concentrado da separação em líquido denso da fração - 1 + 0,150 mm



Concentrado da flotação.

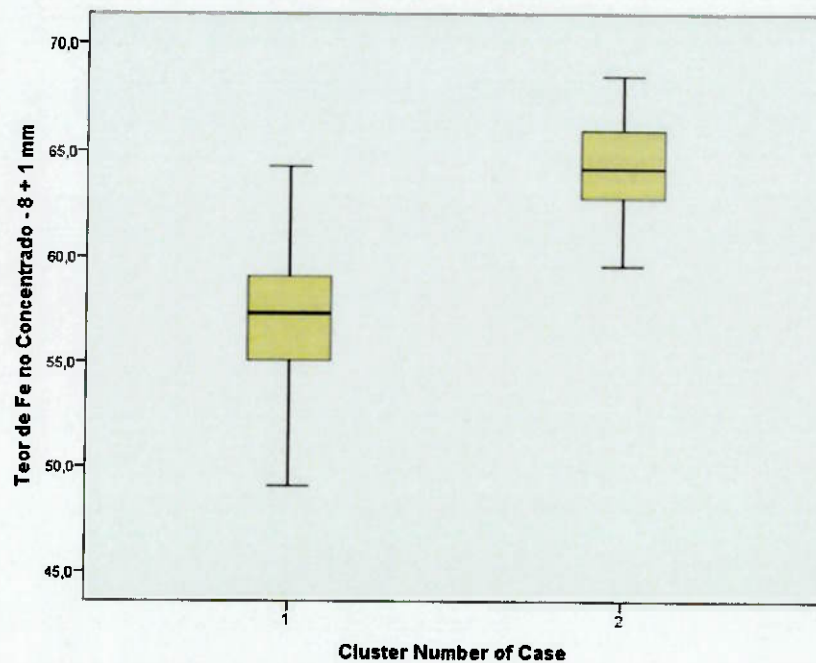


# Dendrograma da análise por agrupamento hierárquico.

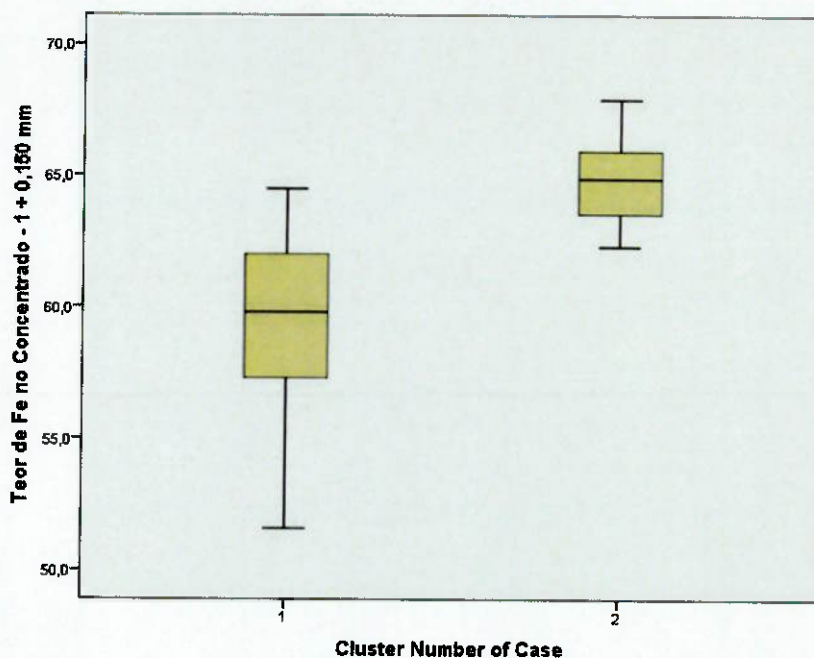


*Boxplots* da análise por agrupamento – Grupos 1 e 2.

Concentrado da separação em líquido denso da fração - 8 + 1 mm.



Concentrado da separação em líquido denso da fração - 1 + 0,15 mm.



### Concentrado da flotação.

