

Universidade Federal do Rio Grande do Sul
Programa de Pós-Graduação em Engenharia de Minas

Gustavo Leandro Tomazi

Erro de Extensão na Amostragem de Minérios

Porto Alegre

2022

Gustavo Leandro Tomazi

Erro de Extensão na Amostragem de Minérios

Dissertação apresentada ao Programa de Pós Graduação em Engenharia de Minas da Universidade Federal do Rio Grande do Sul como parte dos requisitos para obtenção do título de Mestre em Engenharia.

Porto Alegre

2022

Gustavo Leandro Tomazi

Erro de Extensão na Amostragem de Minérios

Esta dissertação de mestrado foi julgada adequada para a obtenção do título de MESTRE EM ENGENHARIA, na área de Amostragem, e aprovada em sua forma final pela professora orientadora e pelo Programa de Pós-Graduação em Engenharia de Minas da Universidade Federal do Rio Grande do Sul.

Porto Alegre, 30 do agosto de 2022.

Profa. Vlória Cristina de Souza
Dra. pela Universidade Federal do Rio
Grande do Sul
orientadora

Banca examinadora:

Profa. Dra. Thays de Souza João Luiz (UNIVESP)

Profa. Dra. Vanessa Cerqueira Koppe (UFRGS)

Profa. Dra. Ana Carolina Chieregati (USP)

CIP - Catalogação na Publicação

Tomazi, Gustavo Leandro
Erro de Extensão na Amostragem de Minérios /
Gustavo Leandro Tomazi. -- 2022.
137 f.
Orientadora: Vlândia Cristina de Souza.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Escola de Engenharia, Programa de
Pós-Graduação em Engenharia de Minas, Metalúrgica e de
Materiais, Porto Alegre, BR-RS, 2022.

1. Erros de Amostragem. 2. Erro de Extensão. 3.
QAQC. 4. Variografia. 5. Heterogeneidade. I. de Souza,
Vlândia Cristina, orient. II. Título.

AGRADECIMENTOS

Primeiramente, agradeço a Deus pelo dom da vida.

Sou grato aos meus pais, Domingos Tomazi e Anália Tomazi, pelo inestimável apoio, incondicional suporte e por manterem sempre claro o valor dos estudos.

Presto gratidão à Pietra Fernandes, minha noiva, que exerceu de paciência e caridade durante toda a jornada de conhecimento que contemplou esse trabalho.

Agradeço à Professora Doutora Vlândia de Souza, que, desde a graduação, me ofereceu oportunidades, nunca hesitou em esforçar-se ao máximo para transmitir seus conhecimentos e, com excelência, orientou esse trabalho de forma singular.

Agradeço ao programa de Pós-Graduação em Engenharia de Minas da Universidade Federal do Rio Grande do Sul pela oportunidade de cursar o Mestrado.

Sou grato à CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior e ao CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo suporte.

Por fim, agradeço a todos os amigos e familiares que me apoiaram na realização do Mestrado Acadêmico.

“A fé e a razão constituem como que as duas asas pelas quais o espírito humano se eleva para a contemplação da verdade.”

São João Paulo II

RESUMO

TOMAZI, G. L. **Erro de Extensão na Amostragem de Minérios**. Dissertação (Mestrado em Tecnologia Mineral) – Programa de Pós-Graduação em Engenharia de Minas, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2022.

A amostragem pode ser definida como processo de retirada de incrementos de um lote que se deseja amostrar. Esse processo tem repercussão em todas as etapas de um empreendimento mineiro. Uma amostragem mal conduzida pode resultar em graves consequências, como inviabilidade de um depósito, perda de eficiência em uma planta de beneficiamento ou multas contratuais. Segundo Gy (1982), a amostragem é uma técnica e uma ciência, por vezes, subestimada.

Além dos erros de amostragem vinculados ao protocolo de preparação das amostras de laboratório, bem como tantos outros muito bem descritos por Gy, é preciso considerar os erros de extensão, que dizem respeito ao erro cometido ao estender o resultado das análises das amostras para todo o lote. O erro de extensão pode ser determinado por meio de um teste variográfico especificamente projetado para um determinado estudo ou contexto (objetivos diversos). Assim, dentre os objetivos propostos desta pesquisa estão explicar o conceito físico deste erro e sua dedução matemática, analisando e discutindo seus desdobramentos, tais como as questões: (a) a fórmula de Gy da variância do erro fundamental é realmente acurada?; (b) como o modelo discreto dos erros de amostragem de Gy se relaciona ao modelo contínuo? e (c) como é possível aplicar o cálculo deste erro para os mais diferentes fins (reconciliação, controle de processo e cláusulas contratuais) e como ele se integra a uma nova proposta de QAQC (*Quality Assurance and Quality Control*)?

Ainda, para exemplificar a aplicabilidade do erro de extensão, foi executado um teste variográfico em uma mineradora de carvão com respeito à amostragem no cliente (termoelétrica) para verificar as incertezas. Os resultados do cliente mostraram erros consideráveis de viés e de extensão para o teor de enxofre em especial. Assim, recomendou-se à empresa mineradora diminuir os intervalos de coleta praticados na amostragem automatizada de correia e à termoelétrica cessar a prática de amostragem em pilhas. Em ambos os casos, e em qualquer outro, é fundamental conhecer a variância intrínseca do material/lote e os erros associados à amostragem para otimizar protocolos.

Palavras-chave: erros de amostragem; erro de extensão; erro fundamental; QAQC; carvão; cinzas; variografia; heterogeneidade; reprodutibilidade.

ABSTRACT

TOMAZI, G. L. **Extension Error in Ore Sampling**. Thesis (Master of Science in Mineral Technology) – Post-graduate Program in Mining Engineering, Federal University of Rio Grande do Sul, Porto Alegre. 2022.

Sampling can be defined the process of taking increments from a lot that is to be sampled. This process has repercussions on all stages of the mine life cycle. A badly conducted sampling can result in serious consequences, such as the unviability of a deposit, loss of efficiency in a processing plant, or contractual fines. According to the Gy (1982), sampling is often an underestimated technique and science.

In addition to sampling errors related to the laboratory sample preparation protocol, as well as many others very well were described by Gy, we must consider extension errors, which errors made when extending the result of sample analyses to the entire lot. The extension error can be determined via a variographic test specifically designed for a particular study or context (various objectives). Thus, among the proposed objectives of this research is to explain the physical concept of this error and its mathematical deduction, analyzing and discussing its unfoldings, such as the questions: (a) is Gy's formula of the variance of the fundamental error really accurate? (b) how does Gy's discrete model of sampling errors relate to the continuous model? and (c) how it is possible to apply the calculation of this error for many different purposes (reconciliation, process control and contractual clauses) and how this all integrates regarding a new proposal for QAQC (Quality Assurance and Quality Control).

Also, to exemplify the applicability of the extension error, a variographic test was performed at a coal mining company in order to verify sampling at the customer (thermal power plant) regarding uncertainties. The customer results showed considerable bias and extension errors for sulfur content in particular. Thus, it was recommended that the mining company decrease the collection intervals practiced in automated belt sampling, and that the thermal power plant cease the practice of pile sampling. In both cases, and any other, it is critical to know the intrinsic variance of the material/lot and the errors associated with sampling to optimize protocols.

Keywords: sampling errors; extension error; fundamental error; QAQC; coal; ash; variography; heterogeneity; reproducibility.

SUMÁRIO

1.	INTRODUÇÃO.....	16
1.1	ESTADO DA ARTE	17
1.2	OBJETIVOS.....	20
1.3	ESTRUTURA DA DISSERTAÇÃO	21
2.	REVISÃO BIBLIOGRÁFICA	23
2.1.	ESTATÍSTICA.....	23
2.1.1.	Estatística Descritiva Univariada.....	24
2.1.1.1.	Medida de Localização.....	26
2.1.1.2.	Medidas de Espalhamento.....	30
2.1.1.3.	Medidas de Forma.....	31
2.1.2.	Estatística Descritiva Bivariada.....	35
2.1.2.1.	Gráfico Q-Q Plot	36
2.1.2.2.	Scatterplot e correlação	36
2.1.3.	Distribuições de Probabilidade.....	38
2.1.3.1.	Distribuição Normal.....	38
2.1.3.2.	Distribuição Binominal	40
2.1.3.3.	Distribuição de Poisson.....	41
2.1.4.	Intervalos de Confiança	42
2.1.5.	Erros	43
2.1.5.1.	Erro absoluto	43
2.1.5.2.	Erro relativo.....	44
2.1.6.	Representatividade.....	44
2.1.6.1.	Precisão e Acurácia	45
2.1.6.2.	Heterogeneidade.....	47
2.2.	TEORIA DA AMOSTRAGEM	50

2.2.1.	Teoria da Amostragem Qualitativa.....	52
2.2.1.1.	Erro de Delimitação do Incremento (IDE).....	52
2.2.1.2.	Erro de Extração do Incremento (IEE).....	54
2.2.1.3.	Erro de Preparação dos Incrementos (IPE)	55
2.2.2.	Teoria da Amostragem Quantitativa.....	57
2.2.2.1.	Erro de Flutuação de Heterogeneidade ou de Seleção Pontual (HFE).....	57
2.2.2.2.	Erro de Segregação e Agrupamento (<i>GSE</i>);.....	58
2.2.2.3.	Erro Fundamental de Amostragem (<i>FSE</i>);.....	60
2.2.3.	Protocolo amostral.....	61
2.3.	GEOESTATÍSTICA.....	64
2.3.1.	Variografia.....	65
2.3.1.1.	Análise de Duplicatas e Efeito Pepita	71
2.3.2.	Erro de Extensão (EE)	73
3.	METODOLOGIA.....	77
3.1.	DEDUÇÃO DA FÓRMULA DA VARIÂNCIA DE EXTENSÃO.....	77
3.2.	PROCEDIMENTO PARA CÁLCULO COMPUTACIONAL DO ERRO DE EXTENSÃO	81
3.3.	TESTE VARIOGRÁFICO	87
3.3.1.	Teste Variográfico Personalizado.....	90
4.	ANÁLISE E DISCUSSÃO DOS RESULTADOS	94
4.1.	RELAÇÃO ENTRE MODELO DISCRETO E CONTÍNUO	94
4.2.	ANÁLISE CRÍTICA DO TESTE VARIOGRÁFICO E IMPLICAÇÕES NO ERRO DE EXTENSÃO	104
5.	ESTUDO DE CASO	113
6.	CONCLUSÃO.....	116
7.	REFERÊNCIAS	118

LISTA DE FIGURAS

Figura 1: Esquema de mediana e quartis.	28
Figura 2: Exemplos de gráficos de dados para diferentes variâncias.	30
Figura 3 : Formatos de histogramas de assimetria negativa, respectivamente, simétrica e positiva para teores de minério.	32
Figura 4: Gráficos esquemáticos para fornecer uma interpretação conceitual de vários valores do coeficiente de correlação linear simples, r	37
Figura 5: Distinção entre precisão e acurácia.	45
Figura 6: Distribuições de processos sem viés e enviesados.	46
Figura 7 : Representação de um fragmento mineralizado com constituição heterogênea.	48
Figura 8: Representação de homogeneidade constitucional e distributiva.	49
Figura 9: Representação de constituintes segregados em quatro camadas.	49
Figura 10: Representação de homogeneidade modular.	50
Figura 11: Representação de distribuição completamente aleatória.	50
Figura 12: Exemplo de Erro de Delimitação em pilha de minério.	53
Figura 13: Exemplo de Erro de Delimitação em um lote unidimensional.	54
Figura 14: Exemplo de Erro de Extração por perda de material em perfuração.	54
Figura 15: Representação de um testemunho de sondagem heterogêneo e irregular partido.	55
Figura 16: (a) Alimentação correta e incorreta em um quarteador tipo rifles, (b) Divisão enviesada com quarteador de cruzeta.	56
Figura 17: Erro de seleção pontual periódico em transportador de correia.	58
Figura 18: Pilha de pó de perfuração com grãos segregados e agrupados.	59
Figura 19: Exemplo de protocolo amostral com hipóteses de análise de duplicatas ou quadruplicatas.	63
Figura 20: Amostrador de rejeito de minério de ferro.	64
Figura 21: Exemplo de contribuição dos erros de amostragem no aumento do efeito pepita do variograma.	73
Figura 22: Zona de influência da amostra tomada sobre transportador de correia.	75
Figura 23: Exemplo de erros de interpolação.	75
Figura 24: Carregando banco de dados.	82
Figura 25: Seleção de tipo de objeto.	82
Figura 26: Determinação das variáveis por coluna.	82
Figura 27: Arranjo dos dados espacialmente.	82

Figura 28: Histograma e sumário estatístico dos dados.	83
Figura 29: Ferramenta de Variograma.	84
Figura 30: Carregamento dos dados para variograma.	84
Figura 31: Parâmetros Variograma.....	84
Figura 32: Modelamento da função variograma.....	85
Figura 33: Arquivo de parâmetros Gammabar.	86
Figura 34: Resultado da variância do EE para 7,5 min.	86
Figura 35: Esquema representando a amostragem praticada em pilhar no cliente.....	91
Figura 36: Fluxograma do teste variográficos personalizado.....	93
Figura 37: Somatório de erros em relação a estimativa de teor na preparação de uma amostra.	95
Figura 38: Esquema entre os erros de amostragem dos modelos discreto e contínuo.....	98
Figura 39: Distribuições de resultados de diferentes amostragens frente a um lote.....	101
Figura 40: Comparativo das variâncias dos pares de duplicatas finais em diferentes protocolos amostrais.....	103

LISTA DE TABELAS

Tabela 1: Resultados da pesquisa bibliométrica.....	18
Tabela 2: Produções acadêmicas variância de extensão.....	19
Tabela 3: Antigas e novas notações para os principais erros de amostragem.	51
Tabela 4: Relação entre parâmetros de um variograma e cronovariograma.	69
Tabela 5: Amostras de transportadora de correia.	77
Tabela 6: Amostras de transportadora de correia em diferentes intervalos.....	79
Tabela 7: Resultados do exemplo do cálculo computacional do erro da variância de extensão.	86
Tabela 8: Matriz de teores colorida 4x4 (lotes 30 min).....	104
Tabela 9: Matriz compositada de teores colorida 2x4 (lotes 1h).....	105
Tabela 10: Matriz de teores colorida para teste de alta frequência 4x4 (lotes 30 min).	109
Tabela 11: Matriz de teores colorida 2x4 (lotes 1 h).....	109
Tabela 12: Comparativo entre matrizes 2x4.....	110
Tabela 13: Matriz de teores 1x4 (lotes 2h).	110
Tabela 14: Teores de cinzas e enxofre no ROM e duplicatas: comparação entre resultados obtidos no teste variográfico e os resultados das amostragens de rotina da mineradora e do cliente.	114
Tabela 15: Variância dos erros por intervalo de coleta e intervalos de teores.	115

LISTA DE GRÁFICOS

Gráfico 1: Exemplo de histograma de frequência absoluta.	25
Gráfico 2: Exemplo de histograma de frequência relativa.	25
Gráfico 3: Exemplo de histograma de frequência acumulada.	26
Gráfico 4: Exemplo de <i>blox plot</i>	29
Gráfico 5: Gráfico de probabilidade lognormal.	35
Gráfico 6: Q-q da distribuição dos 100 valores especiais da variável U versus os valores da variável V.	36
Gráfico 7: Função Densidade de Probabilidade de uma variável aleatória normal padrão.	39
Gráfico 8: Gráfico semivariograma e covariância.	68
Gráfico 9: Exemplo de cronovariograma com elementos destacados.	69
Gráfico 10: Cone de busca do variograma com principais elementos.	70
Gráfico 11: Funções que podem constituir o variograma.	71
Gráfico 12: Exemplo variância de extensão x intervalo de coleta.	87

1. INTRODUÇÃO

Quando observamos contratos comerciais entre empresas mineradoras e clientes, entre as cláusulas com maior importância para aceitação de um lote de minério, estão as relacionadas com limites e variações de teores. Não são incomuns multas e penas decorrentes do não cumprimento dessas cláusulas. Uma grande parcela dos problemas de reconciliação dos teores encontrados pelas partes tem origem em deficiências na metodologia de amostragem, por imperícia ou práticas inadequadas, e no desconhecimento de erros de amostragem, como o erro de extensão.

Apesar deste ser, possivelmente, o maior erro de amostragem, quase sempre, ele é ignorado. Ainda, mesmo quando os erros de protocolo de preparação das amostras são determinados, junto com o erro analítico, estas incertezas não retroalimentam os modelos de reconciliação (Abzalov, 2011a, 2016; Chierigati et al. 2008; Chierigati & Pitard, 2009; Marques & Costa, 2013, 2014a, 2014b). Todos precisam estar cientes de que apenas uma amostragem automatizada sozinha não garante o controle de qualidade (Abzalov, 2016; Mucha & Szuwarzyński, 2004).

Assim, este estudo visa mostrar a importância do erro de extensão: o erro cometido com respeito a toda massa do lote que não é coletada durante a amostragem automática sobre transportadores de correia. Quanto maior o intervalo de tempo entre coletas, maior é este erro. Seu conhecimento é essencial nos processos de reconciliação proativa (Chierigati, 2008), ajuda a projetar e otimizar os circuitos e equipamentos de uma usina de tratamento de minério (Pitard, 1993) e norteia a complementação da Teoria da Amostragem clássica de forma a considerar modelos discretos e contínuos.

A ferramenta escolhida para auxiliar no cumprimento dos objetivos deste estudo foi a variografia. Ela pode, de diversos modos e de forma eficaz, ajudar a prevenir e controlar erros de amostragem. Deveria, portanto, fazer parte da base de análises comuns ao denominado programa de controle e garantia de qualidade da amostragem (QAQC, do inglês “*Quality Assurance and Quality Control*”). Entretanto, a execução de testes variográficos desse gênero, comumente, fica restrita à fase de projeto do circuito de tratamento mineral, limitando os benefícios da aplicação dos seus resultados.

Para demonstrar os impactos destes erros de amostragem (Gy, 1979, 1982, 1992; Pitard, 1993; Abzalov, 2016) em valores de teores, em um estudo de caso, foram realizados testes variográficos em três lotes de carvão de uma mineradora, com auxílio da Estatística clássica (Montgomery & Runger, 2010) e da Geoestatística (Matheron, 1971, 1984; Journel & Huijbregts, 1978), usando os softwares *SGems* e *GSLib* (Deutsch & Journel, 1998).

Os lotes analisados foram rastreados até formarem três pilhas em separado no pátio do cliente. Neste local, o protocolo de amostragem é praticamente manual: as pilhas são misturadas com ajuda de caminhões basculantes e retroescavadeiras, sendo achatadas e quarteadas *in loco*, resultando em pilhas de massas cada vez menores. Como meta, buscou-se identificar as fontes geradoras de muitas contratuais e as incertezas inerentes a este processo de reconciliação mineradora/cliente, visando não somente a média, mas as variâncias dos teores de cinzas e enxofre (Gy, 1979, 1982, 1992; Pitard, 1993, François-Bongarçon, 1998).

Os três lotes/pilhas, reconstituídos com sua massa original, alimentaram separadamente um silo que descarregava o material em um transportador de correia. As amostras foram coletadas no menor intervalo de tempo possível, de modo a “esgotar” o lote. Assim, partes do lote que não são acessíveis, quando dispostas numa pilha, têm agora chance de serem coletadas, e as variâncias em função dos intervalos de tempo puderam ser calculadas. Este é procedimento do teste variográfico, o qual permite conhecer a heterogeneidade interna dos lotes (Pitard, 1993 e Gy, 1998) e os erros de amostragem decorrentes.

1.1 ESTADO DA ARTE

Para que seja possível contextualizar o atual panorama das produções acadêmicas pertinentes ao tema deste trabalho, é necessário um breve estado da arte sobre a temática da variância de extensão. Os resultados desta consulta delimitam as linhas de pesquisa já percorridas e norteiam como os novos trabalhos podem contribuir no desenvolvimento do assunto.

A estratégia inicial para a elaboração da pesquisa foi definir palavras-chave de busca capazes de retratar o tema da dissertação. Levando em conta o tema “impacto do erro da variância de extensão da amostragem em empreendimentos mineiros” e a especificidade desse assunto, os únicos termos de busca elencados foram “variância de extensão” e “erro de extensão”. Em seguida, os bancos de dados de pesquisas científicas foram definidos. Com a

finalidade de encontrar trabalhos acadêmicos nacionais, foi considerada a Biblioteca Digital de Teses e Dissertações (BDTD), que integra os sistemas de informação de teses e dissertações existentes nas instituições de ensino e pesquisa do Brasil, e, para a busca com foco em produções internacionais, foi utilizado o Scopus, maior banco de dados de resumos e citações da literatura com revisão por pares.

Os parâmetros de busca, bem como os resultados quantitativos, podem ser observados na Tabela 1. Após a pesquisa nos bancos de dados, uma triagem através da leitura do título e do resumo foi realizada com o objetivo de refinar o número de trabalhos, restando apenas aqueles que realmente são compatíveis com o tema proposto.

Tabela 1: Resultados da pesquisa bibliométrica.

Temática	Impacto do erro da variância de extensão da amostragem em empreendimentos mineiros.	
Base de dados	Scopus	BDTS
Filtros de busca	Pesquisar em: Palavras-chave, título e resumo; Tipo de documentos: artigos, revisões, breve pesquisa, notas, artigos de conferências e revisões de conferências; Anos: sem filtro; Idiomas: sem filtro.	Pesquisar em: título, autor e assunto; Tipo de documentos: dissertações e teses; Anos: sem filtro; Idiomas: sem filtro.
Código	TITLE-ABS-KEY ("extension variance" OR "extension error")	"variância de extensão" OR "erro de extensão"
Data da busca	15/02/2022	15/02/2022
Total de documentos encontrados	52	2
Documentos após triagem por título e resumo	7	2

Fonte: o Autor (2022).

A próxima etapa foi organizar os resultados e suas informações na Tabela 2.

Tabela 2: Produções acadêmicas variância de extensão.

Ano	Autor(es)	Título
1972	Royle, A.G., Newton, M.J., Sarin, V.K.	<i>Geostatistical Factors in Design of Mine Sampling Programmes.</i>
1978	Sinclair, Alastair J.	<i>Geostatistics and the Earth Sciences.</i>
1996	De Las Cuevas, C.	<i>Use of geostatistical analysis to design sampling programs in salt formations.</i>
1999	Humphreys, M., Shrivastava, P.	<i>Choosing an exploration drillhole spacing: a case study in an iron mine.</i>
2014	Marques, D.M.	Métodos estocásticos aplicados a definição de estratégias de amostragem e homogeneização.
2014	Marques, D.M., Costa, J.F.C.L.	<i>Choosing a proper sampling interval for the ore feeding a processing plant: A geostatistical solution.</i>
2020	Luiz, T.S.J., Souza, V.C.G., Koppe, J.C.	<i>Use of extension variance in monitoring of fluoride in bottled water.</i>
2021	Luiz, T.S.J.	Uso de variância de extensão na amostragem dos parâmetros de águas minerais.
2021	Tomazi, G.L., De Souza, V.C.G.	<i>The most ignored error in the history of mining sampling–extension variance: a case study applied to coal.</i>

Fonte: o Autor (2022).

No total, como resultado refinado da pesquisa nos bancos de dados, 9 produções acadêmicas foram consideradas relevantes. Para análise dos resultados, os trabalhos foram divididos conforme seus períodos de publicação. Na década de 1970, os estudos geoestatísticos davam seus primeiros passos, Royle & Newton & Sarin (1972) introduziram ideias sobre a variância de extensão e demonstraram a influência de parâmetros geoestatísticos nas variâncias. Sinclair (1978), além de apresentar modelos para determinação das variâncias de extensão e dispersão, tentou prever reduções na variância de extensão com o auxílio da Geoestatística.

Após alguns anos sem publicações que incluíssem o tema, De Las Cuevas (1996) focou seu trabalho na minimização das incertezas do teor de água em formações salinas a fim de avaliar intervalos de amostragem otimizados para procedimentos de amostragem representativos. Com uma abordagem voltada para a continuidade espacial, Humphreys & Shrivastava (1999) descreveram o uso de variância de extensão e da variância de estimativa global simplificada para otimizações em grades de perfuração para desmonte. Embora não tenha aparecido nos resultados, no mesmo período, Wackernagel (1998) desempenhou estudos pertinentes à esta dissertação, relacionando os cálculos de variâncias com os suportes envolvidos e examinando amostragens com uma perspectiva Geoestatística.

Após um novo período de tempo sem publicações, Marques & Costa (2014a) e Marques (2014c) investigaram a variografia baseada em simulação dos teores de minério capazes de estimar erros de amostragem. Mais recentemente, Luiz & Souza & Koppe (2020) e Luiz (2021) estimaram os erros relativos na amostragem de água com base no cálculo da variância de extensão. Por fim, Tomazi & Souza (2021) propuseram um teste variográfico em carvão mineral para determinação do erro de extensão, este estudo será apresentado com maiores detalhes no Capítulo 5.

O estado da arte deixou claro que o surgimento das primeiras definições da variância de extensão se entrelaça com o início dos estudos geoestatísticos. No final da década de 90, houveram publicações oferecendo avanços nas concepções matemáticas, tornando os conceitos mais aprimorados. Mas foi só nos últimos anos, talvez resultado de avanços computacionais, que publicações com conteúdo voltado para a implementação da variância de extensão na indústria começaram a surgir. Assim, embora sendo uma temática de grande relevância na amostragem, a variância de extensão não é explorada academicamente conforme o seu grande potencial, sofrendo períodos de estagnação de produções científicas. Com as últimas publicações, esse cenário apresenta sinais de mudança para os próximos anos.

1.2 OBJETIVOS

A seguir são descritos o objetivo geral e os objetivos específicos deste trabalho.

1.2.1 Objetivo Geral

O objetivo geral desta pesquisa é explicar o que é o erro de extensão de amostragem e os impactos consequentes em empreendimentos mineiros. Ainda, analisar como o modelo discreto dos erros de amostragem de Gy se integra ao modelo contínuo que abarca os erros de extensão.

1.2.2 Objetivo Específicos

- Explicar o conceito de erro de extensão e suas implicações no empreendimento mineiro. E, tomando como ponto de partida os protocolos de QAQC, identificar as distinções entre os tipos de erros presentes na amostragem mineral;
- Explanar os fundamentos matemáticos necessários ao cálculo da variância de extensão (relações de Krige e questões de mudança de suporte) e demonstrar a dedução da fórmula principal;
- Elucidar o que é um teste variográfico e suas aplicações, uma vez que é por meio deste que se pode acessar o erro de extensão;
- Expor o que deveria ser realmente um programa de QAQC, onde, para que seja possível obter as estimativas das incertezas, é necessário somar o erro de extensão ao do protocolo de preparação das amostras (incluindo a variância relativa de pares de duplicatas);
- Analisar e discutir as limitações da fórmula do erro fundamental de Gy com ênfase na questão da taxa de redução de massa dentro do objetivo geral de integrar o modelo discreto de Gy ao contínuo (Geoestatística);
- Demonstrar teoricamente porque a fórmula de Gy não é acurada e, portanto, não seria apropriada para otimização de protocolos de preparação de amostras em laboratório e muito menos para o cálculo dos erros totais de amostragem;
- Exemplificar a aplicação destes conceitos com exemplos teóricos e um estudo de caso de carvão mineral.

1.3 ESTRUTURA DA DISSERTAÇÃO

A dissertação encontra-se estruturada da seguinte forma:

Capítulo 1 - Introdução - Importância da amostragem e do erro de extensão para um empreendimento mineiro e desdobramentos na cadeia produtiva;

Capítulo 2 - Revisão Bibliográfica - Estatística básica; teoria da amostragem (conceito de heterogeneidade e modelo discreto de Gy - erro fundamental); Geoestatística básica (cronovariografia e conceito geral do erro de extensão);

Capítulo 3 – Metodologia - Erro de extensão (fundamentos matemáticos, representação física do fenômeno, dedução da fórmula desde a relação de Krige); passo a passo do cálculo do erro de extensão, com base nos algoritmos disponíveis (*open source*); projeto de um teste variográfico, visando (A) reconciliação e (B) e implicações contratuais; relação entre modelo discreto e contínuo dos erros de amostragem; QAQC - uma nova proposta pensando em reconciliação mina e usina;

Capítulo 4 - Análises e discussão - Limitações da fórmula de Gy (ênfase na taxa de redução de massa, vieses intrínsecos) e suas consequências; relação entre modelo discreto e contínuo de amostragem; projeto de teste variográfico - vantagens e limitações; importância da amostragem e do erro de extensão para um empreendimento mineiro e desdobramentos na cadeia produtiva; nova proposta de QAQC;

Capítulo 5 - Estudo de caso aplicado em minério de carvão;

Capítulo 6 - Conclusões e proposta de continuidade do trabalho;

Capítulo 7- Referências Bibliográficas.

2. REVISÃO BIBLIOGRÁFICA

2.1. ESTATÍSTICA

A Estatística possui conceitos particularmente próprios; é relevante que os mais pertinentes para o estudo sejam explanados. Os dados de uma análise Estatística são obtidos de uma população; a população é definida, segundo Guedes (2005), como conjunto de elementos que possuem ao menos uma característica em comum. Já o subconjunto representativo de elementos de uma população, com dimensão menor que a da população, é chamado de amostra, obtido através da amostragem, conceito que será aprofundado no decorrer do estudo. As unidades da população são chamadas de elementos.

Uma vez adquiridos os dados, a característica que será observada, medida ou contada nos elementos da população ou da amostra, que pode adquirir valores diferentes de elemento para elemento, é chamada variável. Dessa forma existe uma distinção entre diferentes tipos de variáveis:

- i. Variável qualitativa: assume como possíveis valores, atributos ou qualidades, também denominada categórica. Podendo ser:
 - a. Variável qualitativa nominal: assume atributos ou qualidades como possíveis valores, e estes não apresentam uma ordem natural de ocorrência. Ex.: cor, sexo;
 - b. Variável qualitativa ordinal: assume atributos ou qualidades como possíveis valores, e estes apresentam uma ordem natural de ocorrência. Ex.: ano escolar, meses do ano.
- ii. Variável quantitativa: é uma variável que assume números como possíveis valores. Podendo ser:
 - a. Variável quantitativa discreta: assume como possíveis valores números, em geral inteiros, formando um conjunto finito ou enumerável. Ex.: número de filhos, número de bactérias por litro de água;

- b. Variável quantitativa contínua: assume como possíveis valores números em uma escala contínua, para os quais valores fracionais fazem sentido. Ex.: peso, tempo, cota.

Ainda, existem as variáveis quantitativas aleatórias, elas podem ser entendidas como resultado de uma função numérica cujo valor é determinado por fatores de chance, ou seja, podem estar sujeitas à influência conjunta dos fatores associados ao experimento que interagem conjuntamente.

2.1.1. Estatística Descritiva Univariada

A Estatística Descritiva tem por objetivo descrever a realidade observada (população ou amostra), usando métodos numéricos e métodos gráficos, realizando comentários simples de maneira informativa. Pode-se dizer que a Estatística Descritiva se ocupa do tratamento de dados quantitativos (ordenação, exposição e sumarização de registros e dados) do fenômeno de estudo. Para Larson & Farber (2010), Estatística Descritiva envolve a organização, o resumo e a representação de dados.

Com a simples observação de um banco de dados, na maioria das vezes, é impossível a identificação de padrões. Uma forma de organizar um conjunto de dados é agrupando-os em intervalos de classes, assim formando uma distribuição de frequências. Dessa forma, o modo mais fácil de visualizar padrões é através de um histograma. Um histograma de frequência é um diagrama de barras que representa a distribuição de frequências de um conjunto de dados. De acordo com Larson & Farber (2010), um histograma possui as seguintes propriedades: escala horizontal (quantitativa e mede os valores dos dados), escala vertical (mede as frequências das classes) e as barras consecutivas, elas devem começar e terminar na fronteira da classe.

As frequências podem ser medidas de forma:

- i. Absoluta: utiliza os valores absolutos de cada classe (Gráfico 1);
- ii. Relativa: utiliza valores proporcionais de cada classe (Gráfico 2);
- iii. Acumulada: adiciona a cada classe anterior, indicando quantos elementos estão abaixo de um certo valor (Gráfico 3).

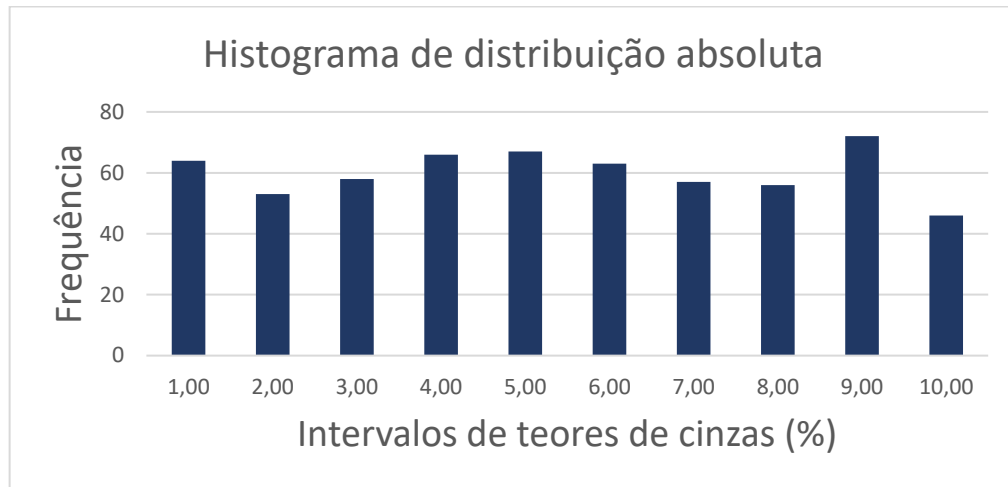


Gráfico 1: Exemplo de histograma de frequência absoluta.

Fonte: o Autor (2022).

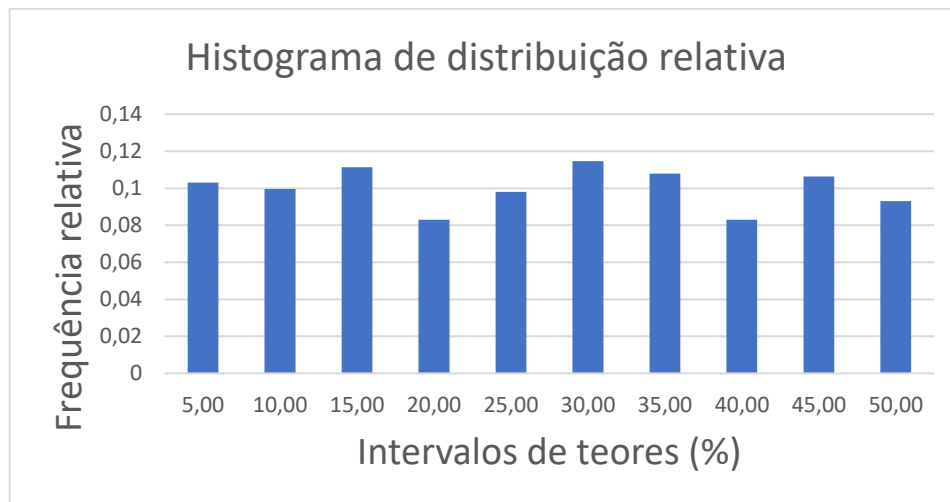


Gráfico 2: Exemplo de histograma de frequência relativa.

Fonte: o Autor (2022).

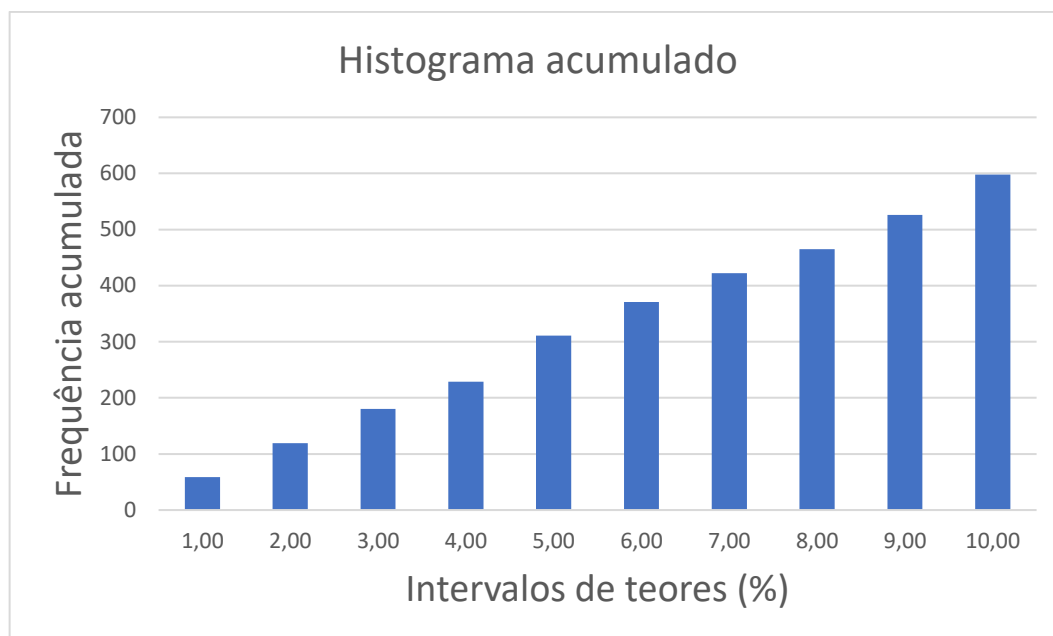


Gráfico 3: Exemplo de histograma de frequência acumulada.

Fonte: o Autor (2022).

As características importantes da maioria dos histogramas podem ser obtidas por um resumo de estatísticas, chamado sumário estatístico. As características resumidas no sumário se dividem em três categorias: medidas de localização, medidas de propagação e medidas de forma.

O sumário estatístico conta com a média, a mediana e a moda, que podem nos indicar onde está o centro da distribuição. A localização de outras partes pertinentes da distribuição é obtida por quantis, estes são usados para descrever a variabilidade dos valores da distribuição. O coeficiente de assimetria descreve a forma da distribuição e o coeficiente de variação fornece informações sobre a duração da cauda em certos tipos de distribuições. Juntas, essas estatísticas fornecem um resumo valioso das informações contidas no histograma.

2.1.1.1. Medida de Localização

Os histogramas de frequência fornecem maior quantidade de informações a respeito das variáveis se comparado com uma simples tabela de dados. Desejando resumir estes dados, apresenta-se três valores que tem maior importância na representatividade na localização da distribuição média, moda e mediana. A medida de localização mais utilizada é a média. A média

é definida como a soma das entradas de dados dividida pelo número total de entradas (Equação 1). Ela é afetada diretamente por valores extremos e levemente afetada por valores centrais.

$$m = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{Equação 1}$$

Onde:

m: média;

n: número de entradas;

x_i : valor de cada entrada.

A mediana é o ponto médio dos valores observados se eles são organizados em ordem crescente. Metade dos valores está abaixo da mediana e metade dos valores acima, calculada conforme Equação 2. A aplicação da mediana requer que a variável possa ser ordenável. Uma desvantagem da mediana é ser muito afetada por ligeiras variações nos valores centrais da variável. Em compensação, ela permanece inalterada para mudanças ainda que grandes nos valores extremos da distribuição.

$$M = \begin{cases} \frac{x_{n+1}}{2} & \text{se } n \text{ é ímpar} \\ (x_{\frac{n}{2}} + x_{\frac{n}{2}+1})/2 & \text{se } n \text{ é par} \end{cases} \quad \text{Equação 2}$$

A moda é o valor que ocorre com mais frequência. A classe com a barra mais alta no histograma dá a ideia de onde ela está localizada. Uma desvantagem da moda é a sua alteração a medida que a precisão dos valores dos dados aumenta. É extremamente sensível a pequenas alterações nos valores observados, não sendo tão útil com números constituídos por vários dígitos significativos (Isaaks & Srivastava, 1989).

Ainda existem outras medidas interessantes, como o mínimo, menor valor no conjunto de dados. Em muitas situações práticas, os menores valores são registrados simplesmente como abaixo de um limite de detecção, em tais situações, pouco importa para fins descritivos. O máximo é o maior valor no conjunto de dados.

Da mesma forma que a mediana divide os dados em metades, os quartis os dividem em quartos (Figura 1). Se os valores dos dados são organizados em ordem crescente, então um quarto dos dados fica abaixo do quartil inferior, ou do primeiro, Q1, e um quarto dos dados cai acima do quartil superior ou terceiro, Q3.

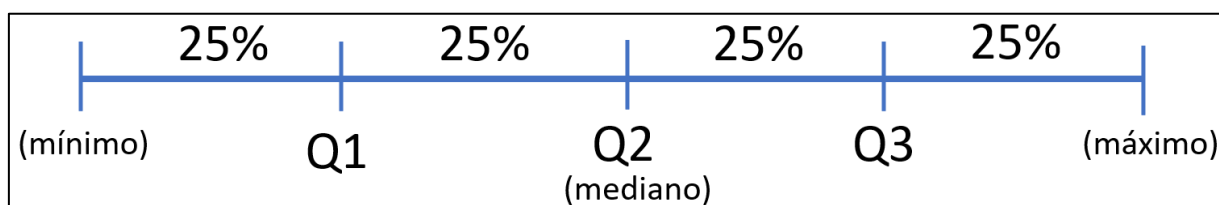


Figura 1: Esquema de mediana e quartis.

Fonte: o Autor (2022).

A ideia de dividir os dados em metades, com a mediana, ou em quartos, com os quartis, pode ser estendida a qualquer fração. Os decis dividem os dados em décimos. Os quantis ($q(p)$), são valores no qual temos no histograma cumulativo igual a $p\%$ (divide a distribuição em duas partes).

Ainda se tratando de localização da distribuição, existe a visualização *box splot* (gráfico de caixa), um gráfico utilizado para avaliar a distribuição empírica dos dados, é formatado com o primeiro e terceiro quartis e a linha da mediana. O retângulo formado contém 50% dos valores do conjunto de dados e a amplitude do gráfico, representado pelas hastes inferior e superior, diz respeito a dispersão dos dados (Gráfico 4).

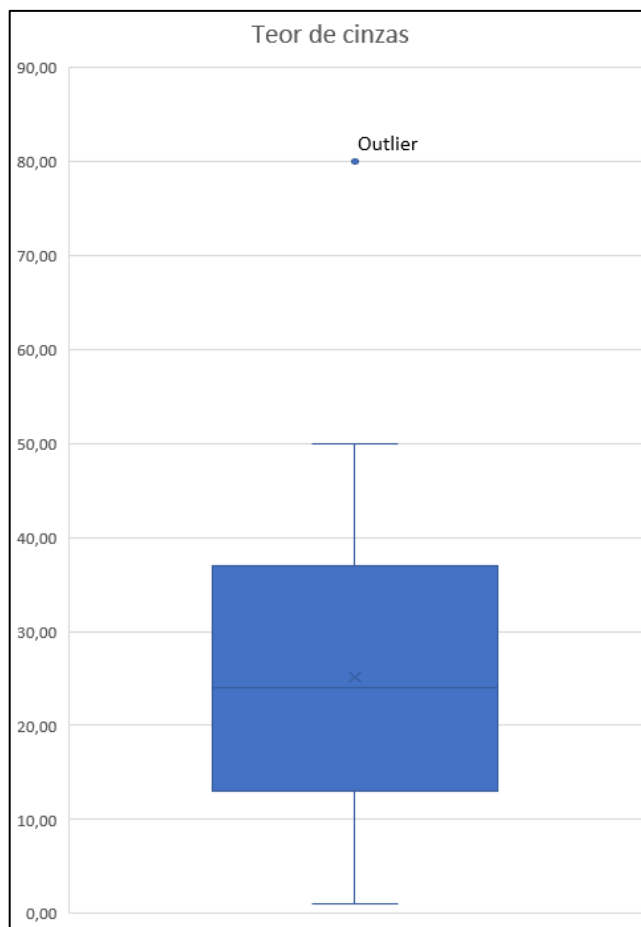


Gráfico 4: Exemplo de *box plot*.

Fonte: o Autor (2022).

É possível observar um ponto com valor elevado em relação a distribuição dos demais no Gráfico 4. Medidas que apresentam essas características destoantes podem ser valores extremos que pertencem à solução do problema e devem ter impacto significativo na estimativa, ou podem ser valores discrepantes da distribuição gaussiana que não são relevantes para a solução da meta imposta pelo estudo. A distinção entre os dois necessita de informações qualitativas, normalmente contidas nas amostras propriamente ditas. Chamados de *outliers*, são definidos como valores que extrapolam a distribuição dos dados, ou seja, a média mais ou menos 3 desvios-padrões.

2.1.1.2. Medidas de Espalhamento

É válido afirmar que quanto maior o espalhamento dos dados, maior será a dificuldade na execução da análise. É conveniente utilizar medidas que sumarizam a variabilidade de um conjunto e que permitam comparar conjuntos diferentes de valores segundo algum critério. Os critérios mais adotados para tal fim levam em conta a dispersão dos dados em torno da sua média, desvio padrão e variância, conforme Morettin & Bussab (2004).

A variância é a medida de espalhamento dos valores amostrais ao redor da média. De forma resumida, se os valores tendem a se concentrar próximos da média, a variância será pequena, e se os valores tendem a se concentrar afastados da média, a variância será grande (Figura 2).

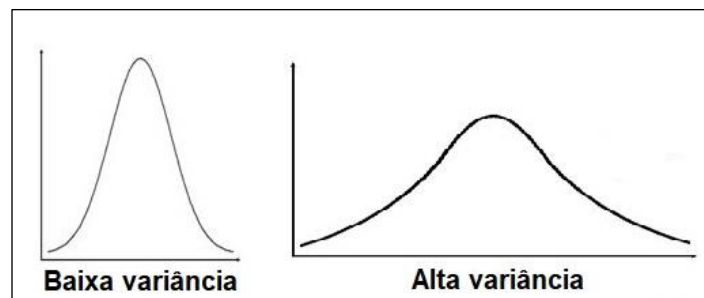


Figura 2: Exemplos de gráficos de dados para diferentes variâncias.

Fonte: o Autor (2020).

A definição de variância consiste na diferença quadrática média entre os dados e a média dos dados, Equação 3). O desvio padrão é simplesmente a raiz quadrada da variância (Equação 4).

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 \quad \text{Equação 3}$$

Onde:

σ^2 : variância;

n: quantidade total de dados;

x_i : valor dos dados;

m: média.

$$\sigma = \sqrt{\sigma^2}$$

Equação 4

Onde:

σ : desvio padrão.

Outra medida útil no estudo da disseminação dos valores observados em um banco de dados é o intervalo interquartil (IQR), definido como a diferença entre o quartil superior (Q3) e o inferior (Q1). Por não usar a média como centro da distribuição, é preferido como medida de espalhamento quando poucos valores extremos influenciam a média.

Além dessas, também podemos trazer aqui a Esperança Matemática $E(x)$, útil em situações que precisamos definir o valor esperado e temos mais de um possível valor para a variável, ou seja, chances diferentes para a ocorrência desses valores. Conforme Yamamoto & Landim (2013), dada a variável aleatória X discreta, com função de probabilidade $P(X=x)$, a esperança matemática é dada conforme Equação 5 e a variância em termos de esperança pela Equação 6.

$$E(x) = \sum_i xP(X = x)$$

Equação 5

$$\sigma^2(x) = \sum_i (x - E(X))^2 P(X = x)$$

Equação 6

2.1.1.3. Medidas de Forma

Na busca da análise da forma de um histograma, é habitual recorrer à medida de assimetria. Partindo das três medidas de tendência central citadas anteriormente, é possível observar o tipo de distribuição que o conjunto de dados apresenta. Ao considerarmos o formato de um histograma de frequências podemos classificá-lo (Figura 3) em:

- i. Distribuição simétrica: Média = Mediana = Moda;
- ii. Distribuição positiva ou para a direita: Moda < Mediana < Média;
- iii. Em distribuições com assimetria negativa ou para a esquerda: Média < Mediana < Moda;

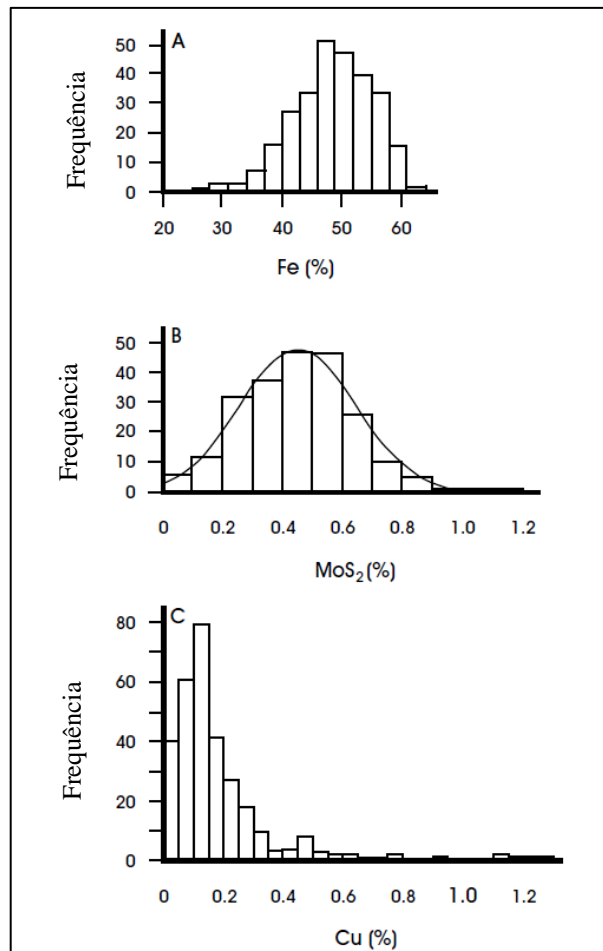


Figura 3 : Formatos de histogramas de assimetria negativa, respectivamente, simétrica e positiva para teores de minério.

Fonte: modificada, Sinclair & Blackwell (2004).

Para verificar a simetria do histograma, também é possível utilizar métodos matemáticos. O coeficiente de assimetria (CS) (Equação 7) verifica a magnitude da assimetria, quanto mais próximo de zero o valor, maior será a simetria da distribuição. O coeficiente de assimetria sofre mais do que média com a sensibilidade a valores altos erráticos. Um único valor grande pode influenciar fortemente o coeficiente de assimetria, pois a diferença entre cada valor de dados e a média é cubada. Muitas vezes, não se usa a magnitude do coeficiente de assimetria, mas apenas o seu sinal para descrever a simetria.

Em um conjunto de dados geoquímicos, por exemplo, uma assimetria positiva é típica quando a variável que está sendo descrita é a concentração de um elemento menor, já uma cauda de pequenos valores à esquerda do histograma, com mediana maior que a média, é típico para concentrações de elementos principais, o histograma será inclinado negativamente, segundo Isaaks & Srivastava (1989).

$$CS = \frac{\frac{1}{n} \sum_{i=1}^n (xi - m)^3}{\sigma^3} \quad \text{Equação 7}$$

Onde:

CS: coeficiente de assimetria;

σ : desvio padrão.

Outra estatística frequentemente usada como alternativa para descrição da distribuição é o coeficiente de variação (CV) (Equação 4). É principalmente adotada para distribuições com todos valores positivos e com assimetria também positiva, embora possa ser calculado para outros tipos de distribuição.

$$CV = \frac{\sigma}{m} \quad \text{Equação 8}$$

Onde:

CV: coeficiente de variação;

Se a estimativa é o objetivo final de um estudo, o coeficiente de variação pode fornecer algum aviso de problemas futuros. Um coeficiente de variação maior que 1 indica a presença de alguma alta errática, há valores amostrais que podem ter um impacto significativo nas estimativas finais.

Por fim, os gráficos de probabilidade são fortes aliados na verificação do comportamento de determinados fenômenos e se agrupamentos de amostras possuem mais de uma população. Os eixos de um gráfico de probabilidade são construídos de forma que a percentagem acumulada se apresente como uma linha reta caso os dados se comportem segundo uma distribuição normal. De maneira análoga, um gráfico de probabilidade lognormal permite testar a hipótese dos dados se comportarem segundo um modelo de distribuição lognormal.

Em ambos os tipos de gráficos, a população é caracterizada pelos pontos se comportarem de maneira retilínea. No caso dos pontos sugerirem uma inflexão ou quebra, podemos estar lidando com duas populações distintas (Gráfico 5), chamado gráfico bimodal. Fisicamente, indica uma anomalia, como um depósito mineral ou uma pluma de contaminação por exemplo. Uma terceira situação seria a presença de *outliers* em certa quantidade, fato que pode representar uma outra população diferente que, por possuir baixa amostragem, os pontos não foram capazes de caracterizar a reta da população.

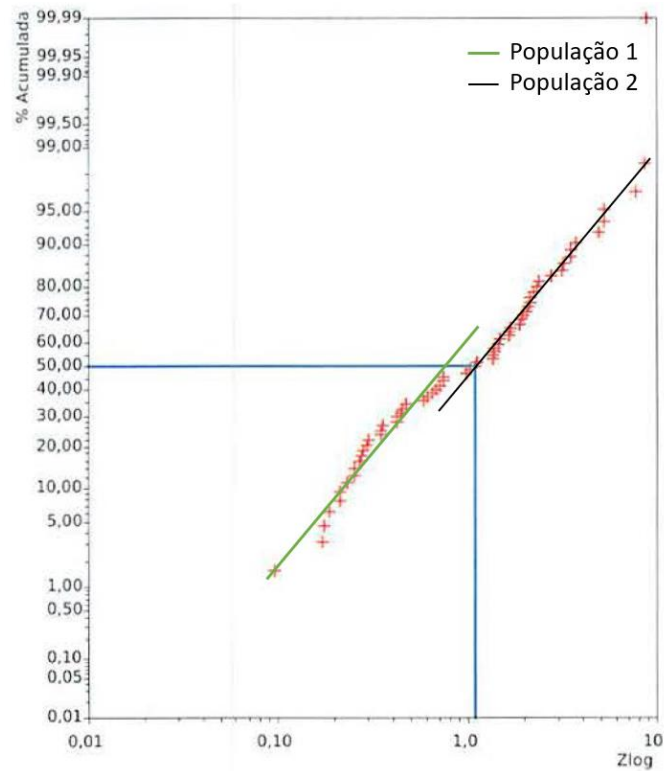


Gráfico 5: Gráfico de probabilidade lognormal.

Fonte: modificado, Yamamoto & Landim (2013).

2.1.2. Estatística Descritiva Bivariada

Algumas das características mais importantes e interessantes da Terra são os conjuntos de dados científicos relacionados com dependência entre variáveis, Isaaks & Srivastava (1989). Mesmo observando, através do sumário estatístico univariado, duas variáveis com forma e espalhamento dos dados diferentes, não é o suficiente para afirmar que as variáveis não possuem dependências entre si. É justamente nesse contexto que entram as ferramentas de descrição bivariadas.

A Estatística Bivariada permite estabelecer relações entre as variáveis, determinando se as diferenças entre as distribuições são estatisticamente significantes, de modo a descobrir o quanto elas são associadas. Seu principal objetivo é ajudar a entender o comportamento de duas variáveis e, se possível, poder estimar uma delas a partir da outra, mediante uma regressão linear.

2.1.2.1. Gráfico Q-Q Plot

Para uma boa comparação visual de duas distribuições, podemos usar um gráfico chamado Q-Q Plot (Gráfico 6). Essa ferramenta visual é comumente utilizada quando há razões para esperar que as distribuições sejam semelhantes. Nesse tipo de gráfico, os quantis de duas distribuições são plotados versus um ao outro. Se as duas distribuições comparadas tiverem o mesmo número de dados, o cálculo dos quantis de cada distribuição não é necessário, bastando classificar os valores dos dados de cada distribuição em ordem crescente e plotar os pares de valores correspondentes.

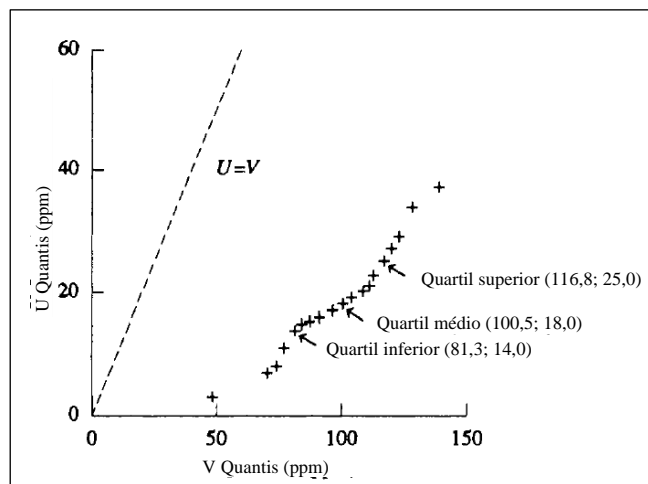


Gráfico 6: Q-q da distribuição dos 100 valores especiais da variável U versus os valores da variável V.

Fonte: modificado, Isaaks & Srivastava (1989).

2.1.2.2. Scatterplot e correlação

A exibição mais comum de dados bivariados é o gráfico de dispersão *Scatterplot*. Nesse gráfico, a coordenada X corresponde ao valor de uma variável e a coordenada Y para o valor da outra variável, a escala pode ser adaptada para uma melhor visualização, até mesmo logarítmica. A análise de *Scatterplot* distingue três casos diferentes de correlação entre as distribuições (Figura 4): positivamente correlacionadas, negativamente correlacionadas ou não possuem correlação.

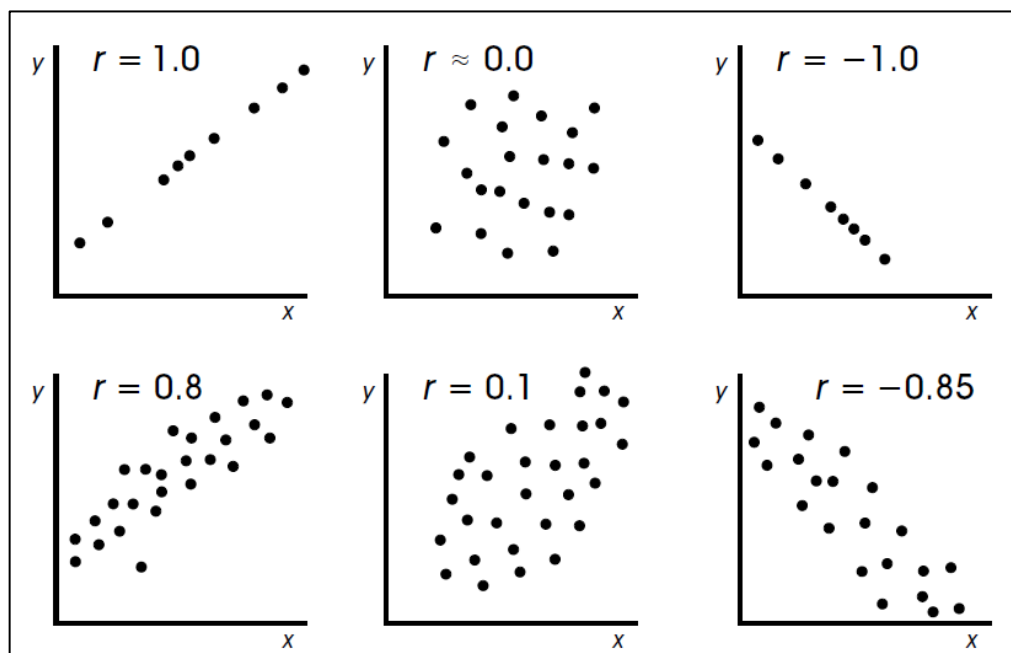


Figura 4: Gráficos esquemáticos para fornecer uma interpretação conceitual de vários valores do coeficiente de correlação linear simples, r .

Fonte: Sinclair & Blackwell (2004).

Duas variáveis serão positivamente correlacionadas se os valores altos de uma variável estiverem associados aos valores altos da outra, em rochas porosas, como a porosidade e permeabilidade apresentam tipicamente correlação positiva em solos. Se os valores altos de uma variável se relacionarem com os valores baixos da outra, possuirão uma correlação negativa. Exemplo disso é a concentração de cálcio e magnésio em dolomitos, estes elementos possuem correlação negativa porque comumente se substituem dentro da rede cristalina dos minerais em função da similitude de raio atômico e valência. No último caso, o aumento dos valores de uma variável não tem efeito aparente na outra, sem correlação (Isaaks & Srivastava, 1989).

Além de fornecer uma boa análise qualitativa de como duas variáveis se relacionam, um *Scatterplot* também é útil ao chamar a atenção para dados destoantes. Nos estágios iniciais do estudo de um conjunto de dados espacialmente contínuo, é necessário verificar e filtrar os dados, pois o sucesso de qualquer método de estimativa depende de dados confiáveis. O gráfico de dispersão pode ser usado para ajudar na validação dos dados iniciais e na compreensão dos resultados posteriores.

Para definirmos a correlação em termos de valores, podemos recorrer ao coeficiente de correlação linear de Pearson $r(x,y)$. Para que o coeficiente não seja influenciado pelas unidades de medidas das variáveis, divide-se a associação linear pelos desvios padrões das variáveis, como demonstra a Equação 9. Segundo Morettin & Bussab (2004), o coeficiente de correlação linear é adimensional e varia de +1 a -1. Próximo ao valor +1 a relação é perfeita e direta, enquanto -1, a relação é perfeitamente inversa e zero é ausência de correlação.

$$r(x, y) = \frac{\sum_i (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_i (x_i - m_x)^2 \sum_i (y_i - m_y)^2}} \quad \text{Equação 9}$$

2.1.3. Distribuições de Probabilidade

A função de distribuição de probabilidade de uma variável aleatória X permite o cálculo da probabilidade de um determinado evento. Segundo Pitard (1993), para descobrir a lei de probabilidade de uma variável aleatória, simplesmente temos que conhecer cada quantidade de probabilidade associada a todos os valores possíveis da variável aleatória. Ainda conforme o mesmo autor, uma lei de probabilidade expressa a totalidade dos vários resultados possíveis de uma variável aleatória associada à sua quantidade de probabilidade correspondente.

Tratando-se de distribuições de variáveis contínuas, a probabilidade de um evento X em um intervalo (a,b) é a área da curva compreendida no intervalo. Em distribuições discretas, em um intervalo (a,b), a probabilidade é a soma das probabilidades dos possíveis valores discretos dentro do intervalo.

2.1.3.1. Distribuição Normal

A distribuição normal tem suas origens no século XIX com Carl Friedrich Gauss em trabalhos sobre erros em observações astronômicas, conforme Morettin & Bussab (2004). É a distribuição de probabilidades mais usada, pois é capaz de modelar diversos processos físicos, biológicos e sociais. Em uma distribuição desse tipo, dizemos que a variável aleatória contínua X precisa apresentar: $-\infty < \text{média} < +\infty$ e $0 < \sigma^2 < +\infty$.

Nesse tipo de função de densidade de probabilidade, a média é onde a distribuição está centralizada, a variância descreve o grau de dispersão e a área total sob a curva do modelo é unitária, tendo como notação $N(m, \sigma^2)$. O Gráfico 7 demonstra essa distribuição e a variável aleatória Z é definida com a Equação 10.

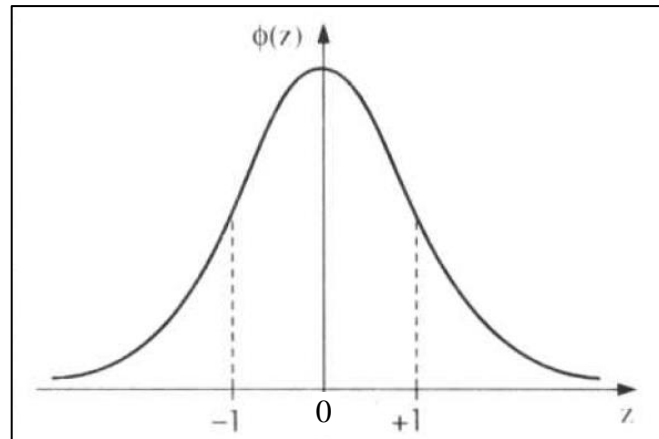


Gráfico 7: Função Densidade de Probabilidade de uma variável aleatória normal padrão.

Fonte: modificada, Morettin & Bussab (2004).

$$Z = \frac{x - m}{\sigma}$$

Equação 10

Conforme Montgomery & Runger (2010), a função densidade de probabilidade está definida na Equação 11. Já o cálculo da probabilidade em um intervalo $P(a < X < b)$ é expresso pela Equação 12, definida como a área abaixo da curva da função. Normalmente, tem seu resultado aproximado por métodos numéricos por não possuir solução analítica.

$$f_x = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

Equação 11

$$P(a < x < b) = \int_b^a \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

Equação 12

Ainda, o gráfico da distribuição normal pode ser interpretado de tal forma que os limites de confiança correspondem:

- 68,3% dos valores individuais quando nos referimos a média ± 1 desvio;
- 95,5% dos valores individuais, quando nos referimos a média ± 2 desvios;
- 99,7% dos valores individuais, quando nos referimos a média ± 3 desvios.

A grande aplicabilidade da distribuição normal tem sua justificativa no Teorema do Limite Central. No teorema, é demonstrado que a distribuição das médias de variáveis com distribuição qualquer (mesmo que a distribuição seja assimétrica) converge para o modelo normal com o aumento de número de elementos. O teorema trabalha com variância relativa, assim, definimos o Coeficiente de Variação (Equação 8) como o desvio padrão relativo (σ_{Rel}) e obtemos a variância relativa (σ_{Rel}^2) na Equação 12.

$$\sigma_{Rel}^2 = \left[\frac{\sigma_{Abs}}{m} \right]^2 \quad \text{Equação 12}$$

2.1.3.2. Distribuição Binominal

A distribuição binomial tem sua aplicação em experimentos repetidos, uma sequência de sucessos e fracassos ou de uns e zeros por exemplo. Os ensaios são independentes nesta distribuição, os resultados não exercem influência um nos outros, existem apenas dois resultados possíveis e a probabilidade $P(x=k)$ é constante em cada ensaio.

Ainda, segundo Pitard (1993), dois fragmentos totalmente liberados seguem uma distribuição binomial quando misturados. A probabilidade dessa distribuição pode ser definida com a Equação 13. Para o autor, esta distribuição tem muitas aplicações práticas e levará naturalmente ao modelo de Poisson para elementos traço, quando q se torna muito grande e p muito pequeno.

$$P(x = r) = C_n^r p^r q^{n-r} \quad \text{Equação 13}$$

Onde:

C_n^r : combinação de r objetos de um conjunto que contém n objetos;

p : chance de seleção do componente de interesse;

q : chance de seleção de outros componentes;

n : número de tentativas;

r : número de sucessos.

2.1.3.3. Distribuição de Poisson

Pode ser considerada como a distribuição discreta associada a eventos raros. Conforme Pitard (1993), esse tipo de distribuição geralmente ocorre quando um componente está presente em um lote no nível de traço, ou seja, é um caso limite do modelo binomial. Segundo o autor, à medida que as amostras se tornam muito pequenas, a probabilidade de ter um elemento traço presente em uma amostra selecionada diminui drasticamente. Quando a Equação 13 é aplicada em uma situação que n é muito grande e a chance p de selecionar um grão é muito pequena, a probabilidade para dois valores sucessivos de x é descrita pelas Equação 14 e Equação 15.

$$P(r) = \frac{n!}{(n-r)! r!} p^r q^{n-r} \quad \text{Equação 14}$$

$$P(r-1) = \frac{n!}{(n-1)! (n-r+1)!} p^{r-1} q^{n-r+1} \quad \text{Equação 15}$$

Aplicando a razão $P(r)/P(r-1)$, multiplicando cada termo da equação por n , considerando n sendo muito grande, $q \gg 1$ e r pequeno, após algumas transformações matemáticas, obtemos a igualdade $pn = \lambda$. O termo λ pode ser definido, em um contexto de amostragem de particulado, como o número médio hipotético de grãos por amostra.

Com $P(r)/P(r-1) = \lambda/r$, escrevermos sucessivamente $r=1, r=2, \dots, r=x$ e observamos que podemos generalizar essa equação. Assim, aplicando algumas transformações matemáticas, obtemos a probabilidade segundo a Equação 16.

$$P(x = r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad \text{Equação 16}$$

2.1.4. Intervalos de Confiança

O intervalo de confiança (IC) é uma estimativa do intervalo no qual o valor de um parâmetro se encontra com determinada confiabilidade, ou seja, é a probabilidade que esse valor é encontrado dentro desse intervalo. Tomando um exemplo de um experimento com um IC95%, afirmamos que a cada 100 realizações do experimento, o resultado estará dentro do intervalo em 95 vezes.

Assim, a fórmula para o cálculo de um intervalo de confiança pode ser definida segundo a Equação 17 (Morettin & Bussab, 2004).

$$\bar{x} \pm Z \frac{\sigma}{\sqrt{n}} \quad \text{Equação 17}$$

Onde:

Z: valor associado a probabilidade desejada obtido pela distribuição normal;

\bar{x} : média amostral.

Para obtermos Z, é necessário subtrairmos 1-IC, multiplicarmos por 2 e consultarmos sua correspondência na Tabela Z normal acumulada, elaborada através da distribuição gaussiana. Ainda, a fórmula pode ser expressa da seguinte maneira (Morettin & Bussab, 2004).

$$P\left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \quad \text{Equação 18}$$

Onde:

$Z_{\alpha/2}$: valor associado ao intervalo de confiança escolhido obtido pela distribuição normal;

α : nível de significância (1-IC).

Devemos dividir o nível de significância por 2, pois trata-se de uma distribuição bicaudal, e consultar sua correspondência na Tabela Z normal padrão para obtermos o valor $Z_{\alpha/2}$.

2.1.5. Erros

Em diversas áreas científicas que se utilizam de estimativas, ocorre geração de erros, uma vez que, segundo Sinclair & Blackwell (2004), os valores verdadeiros não são perfeitamente conhecidos. Os erros podem ter origem nas mais diversas fontes e a sua quantificação é uma questão fundamental. Para tanto, precisamos definir medidas de erros (ou de exatidão).

2.1.5.1. Erro absoluto

De forma simples, o erro absoluto é a diferença entre o valor medido e o “considerado verdadeiro” (referência adotada em determinado contexto, pois o valor verdadeiro dificilmente é conhecido) em módulo. Conforme demonstrado na Equação 19.

$$E_a = |A - a| \quad \text{Equação 19}$$

Onde:

E_a : Erro absoluto;

A: Valor medido (teor ou granulometria, por exemplo);

a: Valor de referência estabelecido dentro de um contexto.

No caso da amostragem na mineração, nunca obteremos o valor verdadeiro do teor de um lote/depósito por exemplo. Ainda, deve ser considerado um valor de referência para o erro máximo admitido em função de um contexto. Assim, o valor de erro máximo pode ser definido em função da: (a) de reconciliação mina/usina, (b) controle de processo na usina, ou (c) simplesmente um valor contratual estabelecido pelo cliente.

2.1.5.2. Erro relativo

O erro relativo (E_r) é a razão entre o erro absoluto e o valor de referência (A) (Equação 20). Através de uma análise, verifica-se que o erro absoluto se exprime em unidades de grandeza condizentes com o cálculo, enquanto o erro relativo é uma grandeza adimensional, podendo ser multiplicado por 100 para obtenção do percentual.

$$E_r = \frac{E_a}{A} \quad \text{Equação 20}$$

2.1.6. Representatividade

Uma população de onde é retirada uma amostra, no contexto da mineração, pode ser: um depósito mineral inteiro, uma bancada ou um período de tempo de uma correia transportadora por exemplo. Essa população é chamada de lote por Gy (1982), ou seja, o conjunto de unidades que será analisado, e a fração retirada é chamada de incremento, uma amostra pode ser formada por mais de um incremento. Diferentes tipos de lotes, com grandezas e escalas particulares, configuram suportes amostrais desiguais, é necessário cautela ao trabalharmos com distintos suportes comparativamente.

Para que uma amostra de um lote seja considerada representativa, deve ser obtida respeitando a Lei da Equiprobabilidade, conforme Pitard (1993). Uma amostra é equiprobabilística quando a amostragem é probabilística, todos os constituintes têm a probabilidade de serem selecionados diferente de zero, e correta, quando todos tem a mesma probabilidade. Para Gy (2004a), uma alíquota selecionada que não cumpre essa lei é chamada de espécime.

Além disso, quando amostramos, gostaríamos que a população, constituída pelas amostras coletadas, tivesse a mesma média e variância do lote que queremos representar (depósito, bancada, fluxo em correia), isto é, fosse acurada e precisa, assim dizendo, fosse representativa. Entretanto, principalmente em materiais de origem natural, é comum que sua constituição não seja intrinsecamente homogênea, fato que favorece erros de amostragem que nos afastam da representatividade e conduzem a vieses. Assim, para evitar vieses, as definições de acurácia, precisão e heterogeneidade se tornam fundamentais.

2.1.6.1. Precisão e Acurácia

Os erros de amostragem são tradicionalmente determinados em termos de precisão e acurácia dos dados segundo Abzalov (2011b). De maneira simples, a precisão, ou reprodutibilidade, é a medida de o quão próximo os valores das amostras estão próximos um dos outros, e a acurácia é a medida de o quão próximo os valores das amostras estão do valor de referência.

Na Figura 5, estão esquematizadas as distinções entre as distribuições de probabilidade em contextos de acurácia e precisão variados. O centro de um alvo configura o valor referência. Observando as situações (a) baixas precisão e acurácia, (b) baixa acurácia e alta precisão, (c) alta acurácia e baixa precisão e (d) altas precisão e acurácia, concluímos que medidas precisas tem menor dispersão e medidas acuradas se aproximam mais do alvo.

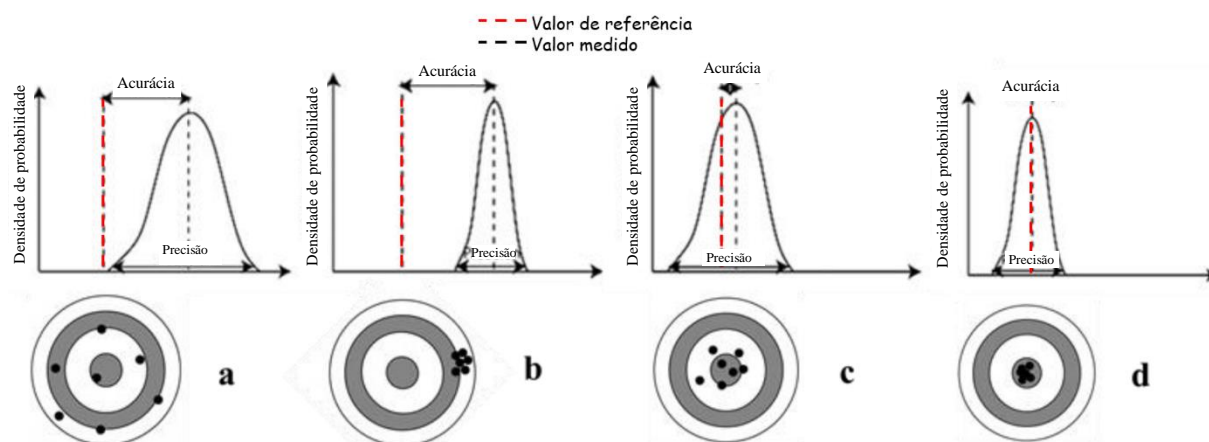


Figura 5: Distinção entre precisão e acurácia.

Fonte: modificada, Zanon (2016).

À medida que as etapas de uma amostragem mineral se sucedem, a massa coletada no lote diminui em processos de divisão em virtude das restrições da quantidade de massa compatível com análise laboratorial. Na Figura 6, observamos o comportamento da distribuição dos valores medidos nas amostras coletadas à medida que as massas reduzem, primeiramente em um processo sem viés, em seguida, em um processo enviesado. Quando não existe viés na preparação amostral, a média aE do parâmetro analisado nas amostras é igual a do lote aL em todas as etapas. Na segunda situação, existe um distanciamento crescente entre as médias medidas bE e a de referência aL ao passo que as etapas se seguem, caracterizando o viés.

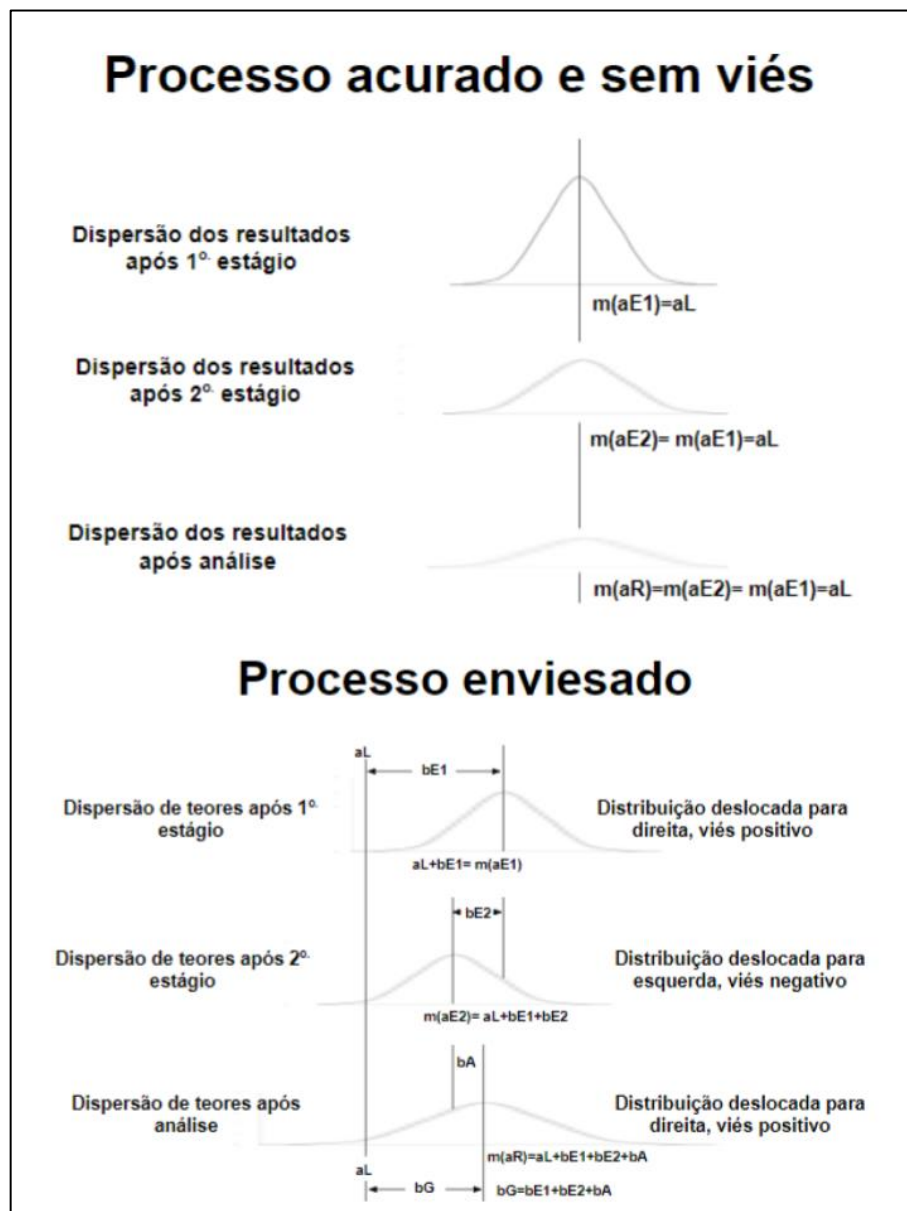


Figura 6: Distribuições de processos sem viés e enviesados.

Fonte: Souza; Costa; Souza (2021).

Dessa forma, uma amostra é representativa quando é tanto acurada quanto precisa, ou seja, a variância do Erro Total de Amostragem Relativo $r^2(TSE)$, definido pela Equação 21, é reduzida ao máximo.

$$r^2(TSE) = m^2(TSE) + \sigma^2(TSE) \quad \text{Equação 21}$$

Onde:

$m^2(TSE)$: erro de acurácia (diferença entre o valor da média medida da população e um valor de referência);

$\sigma^2(TSE)$: erro em termos de variância (diferença entre o valor de variância medida e um valor de referência máximo estabelecido em função de um contexto).

Assim, o erro de precisão será a variância (medida) menos a variância que corresponde a variância intrínseca do minério mais a variância resultante dos erros de amostragem. Além do mais, estes erros só ocorrem justamente pelo minério ser heterogêneo. Nesse sentido, Gy (1998) afirma que jamais devemos nos referir a uma amostra como totalmente precisa, apenas usar o termo se referindo a instrumentos.

2.1.6.2. Heterogeneidade

Um material pode apresentar diferenças intrínsecas próprias de sua natureza em distintas escalas e propriedades, até mesmo o ar parece ser um material homogêneo, quando na verdade não o é, sendo constituído por átomos de Oxigênio, Nitrogênio, entre outros elementos. Um material completamente homogêneo só pode ser definido teoricamente, nunca observado na prática.

Gy (2004b) define a Equação 22 como a forma de calcular a heterogeneidade intrínseca na amostragem mineral considerando os teores dos minerais de valor.

$$h_i = \frac{(a_i - a_L)M_i N_f}{a_L M_L} \quad \text{Equação 22}$$

Onde:

h_i : heterogeneidade intrínseca do incremento;

a_i : teor no incremento;

a_L : teor do lote;

M_i : massa do incremento;

M_L : massa do lote;

N_f : número de fragmentos no incremento.

A confiabilidade nos resultados analíticos depende do controle sobre todos os componentes dos erros de amostragem. Esses componentes surgem a partir da heterogeneidade do material, no caso deste estudo, heterogeneidade dos minérios. A heterogeneidade dos lotes pode ser dividida conforme 2 classificações.

A Heterogeneidade Constitucional do Lote (CH_L), segundo Gy (2004b), é a diferença das propriedades intrínsecas de um conjunto de constituintes, isto é, quando os elementos desse conjunto não têm composição estritamente idêntica (Figura 7). Em sólidos, essa propriedade pode ser alterada com processos físicos de cominuição, como britagem e moagem. Procedimentos de quarteamento e homogeneização não são capazes de alterá-la. A CH_L é definida matematicamente por Gy (2004b) como a variância de h_i na população do lote, $\sigma^2(h_i)$.

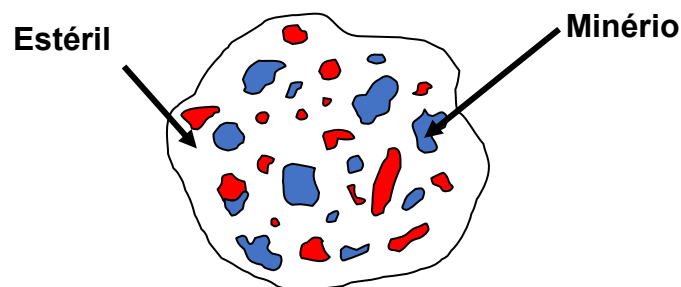


Figura 7 : Representação de um fragmento mineralizado com constituição heterogênea.

Fonte: Souza; Costa; Souza (2021).

A Heterogeneidade Distributiva do Lote (DH_L) é percebida ao observarmos vários grupos de unidades vizinhas, como fragmentos dentro do domínio de um determinado lote. Segundo Pitard (1993), a diferença entre esses grupos de vizinhos é a heterogeneidade distributiva. O oposto, homogeneidade distribucional, ocorre quando todos os grupos de um determinado tamanho podem ser selecionados dentro do domínio do lote. Para Gy (2004b), heterogeneidade distributiva pode ser aumentada por segregação natural ou gravitacional e amenizada com mistura e blendagem.

Tendo em mente “ n ” como sendo um grupo de elementos vizinhos no domínio do lote, a DH_L é definida matematicamente por Gy (2004b) como a variância de h_n na população dos grupos no lote, $\sigma^2(h_n)$.

Considerando uma população composta por quatro tipo de constituintes: círculos, quadrados, diamantes e cruzes. Cada símbolo representa um fragmento sólido quando projetado em um gráfico bidimensional. A Figura 8 e Figura 9 ilustram um material de constituição homogênea e um material segregado em camadas, respectivamente. A Figura 10, para Gy (2004b), representa a única forma de homogeneidade que pode ser encontrada na natureza, homogeneidade modular. Ela é observada onde a mesma sequência unitária de átomos parece se repetir indefinidamente. E, por fim, a Figura 11 mostra uma distribuição completamente aleatória com heterogeneidade distributiva reduzida ao mínimo.

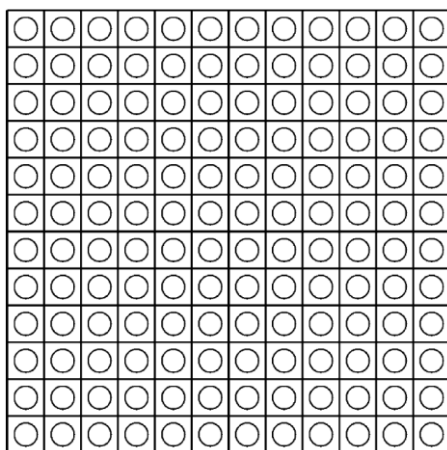


Figura 8: Representação de homogeneidade constitucional e distributiva.

Fonte: Gy (2004b).

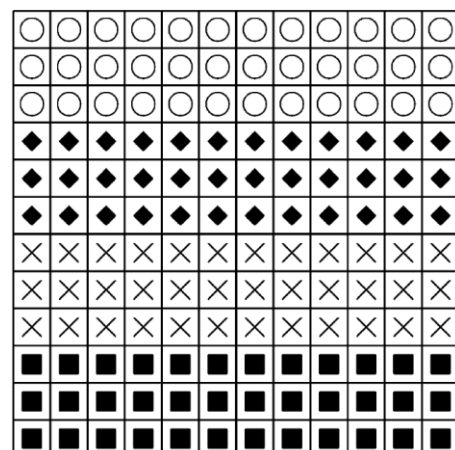


Figura 9: Representação de constituintes segregados em quatro camadas.

Fonte: Gy (2004b).

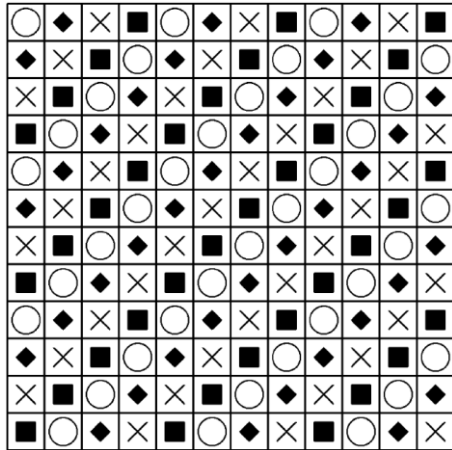


Figura 10: Representação de homogeneidade modular.

Fonte: Gy (2004b).

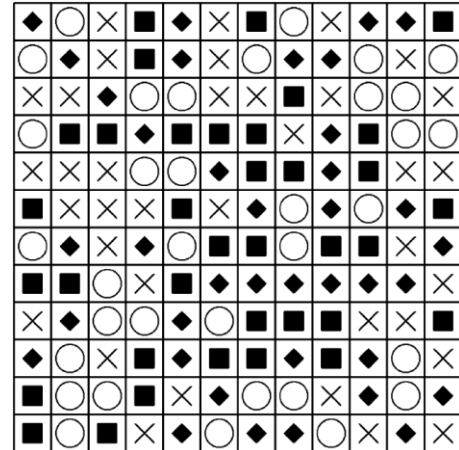


Figura 11: Representação de distribuição completamente aleatória.

Fonte: Gy (2004b).

2.2. TEORIA DA AMOSTRAGEM

A Teoria da Amostragem de Pierre Gy teve o início do seu desenvolvimento na década de 1950, sofreu uma considerável evolução e rege os principais procedimentos de amostragem até hoje. Conforme Gy (2004a), o Erro Global de Estimativa (GEE ou OEE, *Overall Estimation Error*) é constituído por duas componentes, o Erro de Análise (TAE) é a primeira, geralmente representa uma cota menor no resultado da Equação 23 e é associado ao contexto de análise laboratorial química. A segunda fração tem maior influência nos erros de um processo de estimação, é onde estão reunidas as maiores fontes de incertezas e, quando não adulterado, é uma variável aleatória, trata-se do Erro Total de Amostragem (TSE).

$$GEE = TAE + TSE$$

Equação 23

De maneira resumida, Gy (2004a) define a amostragem como uma redução de massa ao mesmo tempo que exalta a sua importância no processo de controle de qualidade. Esse processo tem repercussão em todas as etapas de um empreendimento mineiro. Uma amostragem mal conduzida pode resultar em graves consequências, como inviabilidade de um depósito, perda de eficiência em uma planta de beneficiamento ou multas contratuais.

Ainda a respeito da importância da amostragem na mineração, o processo de reconciliação tem seus resultados diretamente relacionados com a qualidade da amostragem. Segundo Chierigati et al. (2007), a reconciliação pode ser definida como a comparação entre uma estimativa e uma medição, ou seja, entre os teores de minério estimados pelos modelos da jazida e os teores produzidos na usina de beneficiamento. Se a amostragem executada em ambos os contextos não seguir uma metodologia adequada, podem ser criados vieses nos resultados amostrados culminando em tomadas de decisões de planejamento errôneas.

Métodos que priorizam apenas o baixo custo não costumam obter amostras representativas, a economia ocorre apenas em um curto prazo, enquanto prejuízos serão notados em maior escala e a longo prazo. Mediante a esse tipo de problemática, Pierre Gy criou a Teoria da Amostragem (TOS), que zela pela obtenção de uma alíquota representativa do lote, pela fidelidade dos resultados analíticos e diminuição máxima dos erros.

A teoria criada por Gy define o Erro Total de Amostragem (TSE) como o somatório de dois componentes (Equação 24), o primeiro diz respeito ao método e ao uso de equipamentos de coleta e preparo (probabilidade de seleção de incremento não uniforme), chamado Erro de Amostragem Incorreta (ISE). O outro diz respeito a heterogeneidade do material (probabilidade uniforme), Erro de Amostragem Correta (CSE), Gy (1998).

$$TSE = ISE + CSE$$

Equação 24

Os erros de amostragem receberam diversas notações ao longo do tempo desde a criação da Teoria da Amostragem. Segundo Chierigati (2018), em 2009 novas notações foram apresentadas durante a quarta conferência mundial de amostragem e blendagem (World Conference on Sampling and Blending), como apresentado na Tabela 3. Ao longo desta dissertação, as diferentes notações foram utilizadas.

Tabela 3: Antigas e novas notações para os principais erros de amostragem.

Nome do erro	Antiga Notação	Nova Notação
Erro Global de Estimativa (<i>Overall Estimation Error</i>)	OE	OEE
Erro Total de Amostragem (<i>Total Sampling Error</i>)	TE	TSE

Erro Fundamental de Amostragem (<i>Fundamental Sampling Error</i>)	FE	FSE
Erro de Segregação e Agrupamento (<i>Grouping and Segregation Error</i>)	GE	GSE
Erro de Flutuação de Heterogeneidade (<i>Heterogeneity Fluctuation Error</i>)	CE/IE/PSE	FE
Erro de Flutuação de Qualidade (<i>Quality Fluctuation Error</i>)	QE	QFE
Erro de Ponderação do Incremento (<i>Increment Weighting Error</i>)	WE	IWE
Erro de Materialização do Incremento (<i>Increment Materialization Error</i>)	ME	IME
Erro de Delimitação do Incremento (<i>Increment Delimitation Error</i>)	DE	IDE
Erro de Extração do Incremento (<i>Increment Extraction Error</i>)	EE	IEE
Erro de Preparação do Incremento (<i>Increment Preparation Error</i>)	PE	IPE
Efeito Pepita <i>In Situ</i> (<i>In Situ Nugget Effect</i>)	NE	INE

Fonte: Pitard, 2009 apud Chierigati & Pitard, 2018.

2.2.1. Teoria da Amostragem Qualitativa

A Teoria da Amostragem Qualitativa tem a atenção voltada a garantir que as amostras sejam tomadas de forma que todas as alíquotas do lote tenham a mesma chance de serem selecionadas, ou seja, que a Lei da Equiprobabilidade não seja ferida. Quando essa situação não ocorre, nos deparamos com uma amostragem incorreta.

O ISE tem origem em práticas falhas e utilização de equipamentos equivocadamente na tomada e preparação das amostras, são chamados de Erros de Materialização (ME), geralmente sua determinação quantitativa não é acessível. Em suas pesquisas, empiricamente, Gy (2004a) definiu os principais erros que podem compor o ME.

2.2.1.1. Erro de Delimitação do Incremento (IDE)

A depender do contexto onde é realizada a tomada do incremento, surgem diferentes problemáticas para uma delimitação de amostra equiprobabilística, desde o acesso a todas as porções do lote, *desing* de amostradores até a imperícia do profissional responsável, essas e outras variáveis da mesma natureza originam o IDE.

Considerando o exemplo de Abzalov (2011b), uma pilha de minério amostrada com a utilização de uma pá arredondada de coleta (Figura 12): esse instrumento não é capaz de cortar toda a seção devido ao seu formato, deixando sistematicamente um material não amostrado. Este material pode gerar um viés, uma vez que, como ocorre estratificação na pilha, os materiais mais grosseiros e/ou mais densos que se encontram na base não são amostrados

por completo. Ainda, caso a curva granulométrica do material inclua grânulos maiores que o próprio equipamento, também teriam sua probabilidade de seleção zerada.

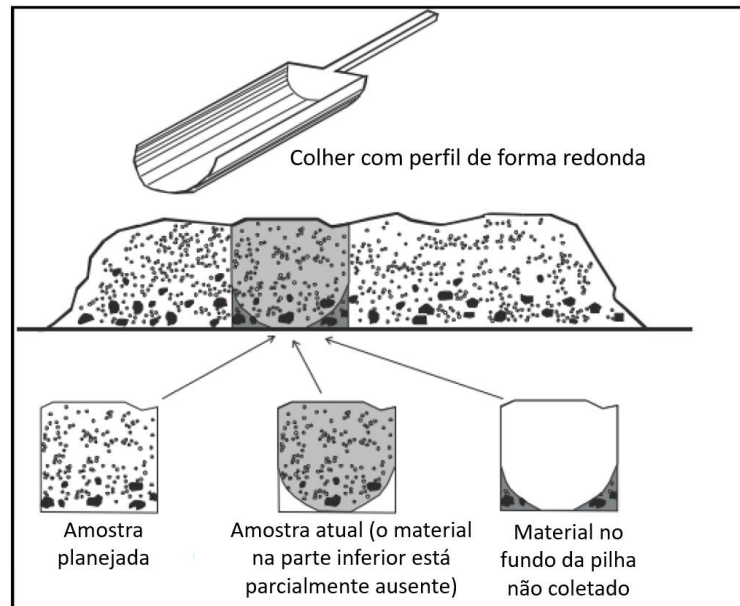


Figura 12: Exemplo de Erro de Delimitação em pilha de minério.

Fonte: modificada, Abzalov (2011b).

Em outra situação, na tomada de amostras em uma correia transportadora, o movimento do dispositivo de amostragem delimita os limites geométricos do domínio espacial ocupado pelos incrementos. Nessa circunstância, os amostradores mais comuns são sobre correia (*Cross Belt Sampler*) e de trajetória retilínea (*Cutter Sampler*). Caso o equipamento seja posicionado em local inadequado, com geometria mal planejada ou velocidade de corte não constante, novamente, o incremento selecionado não será representativo do fluxo de material, podendo acarretar em vieses.

A Figura 13 ilustra tomada de apenas uma fração do fluxo para uma fração do tempo por um amostrador automático de correia, caracterizando uma amostragem não probabilística e enviesada.

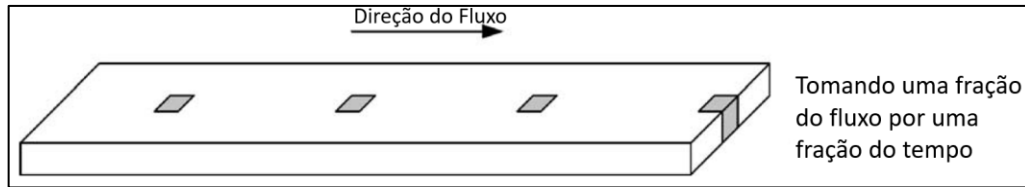


Figura 13: Exemplo de Erro de Delimitação em um lote unidimensional.

Fonte: modificada, Gy (2004a).

2.2.1.2. Erro de Extração do Incremento (IEE)

Este erro, também chamado de erro de recuperação (Pitard, 1993), ocorre quando a amostra não pode ser obtida ou recuperada de maneira representativa. Um exemplo comum deste erro ocorre em amostras de sonda, em que o uso de equipamentos inadequados, ou mesmo da operação incorreta, pode ocasionar perdas de recuperação, privilegiando a amostragem de determinadas partículas e penalizando outras (perda de finos em sondagem), consequentemente tendenciando a amostra. Na Figura 13, é possível observar material particulado fino carregado pelo vento em elevada quantidade durante uma perfuração para desmonte, ou seja, essa porção tem zero probabilidade de ser amostrada.



Figura 14: Exemplo de Erro de Extração por perda de material em perfuração.

Fonte: Souza; Costa; Souza (2021).

Na amostragem de testemunhos de sondagem, além dos problemas que uma baixa recuperação pode causar nas estimativas, fragmentos de minério localizados nas faces do

testemunho podem desprenderem e perderem-se, influenciando nos resultados finais das análises e gerando IEE, Abzalov (2016). Ainda, a tomada de metade do testemunho para o processo de amostragem também gera vieses, pois o cilindro a ser fracionado não é perfeito e o material constituinte é heterogêneo, por vezes, apresentando direções de continuidade preferencias (Figura 15).



Figura 15: Representação de um testemunho de sondagem heterogêneo e irregular partido.

Fonte: o Autor (2022).

2.2.1.3. Erro de Preparação dos Incrementos (IPE)

Logo após a tomada do incremento do lote, uma série de procedimentos são necessários antes que a análise laboratorial seja efetivamente realizada. Conforme Gy (2004a), IPEs podem ocorrer em qualquer um dos estágios: transferência dos incrementos do cortador para o sistema de transporte, transferência dos incrementos para o próximo estágio de amostragem,

identificação, estágios de cominuição, secagem e divisão. O autor separa este erro nas seguintes categorias:

- Contaminação de incrementos e amostras por material estranho;
- Perda de material pertencente ao incremento ou amostra;
- Alteração na composição química;
- Alteração na composição física (umidade e distribuição de tamanho);
- Falhas involuntárias: profissionais despreparados, mesmo agindo de “boa fé”, estão propensos a cometer erros por ignorância, falta de treinamento ou negligência;
- Falhas deliberadas: cometidas por funcionários ou qualquer pessoa que tenha acesso às amostras com objetivo de favorecimento comercial.

Na Figura 16, é possível apreciar dois processos de divisão com distintas fontes de erros de preparação. Em (a), o perfil de um quarteador tipo rifles, alimentado corretamente no seu centro e, ao lado, uma alimentação desigual, com maior quantidade de massa em um dos lados. Tal abordagem leva à uma distribuição desproporcional dos fragmentos e a segregação de partículas mais pesadas de um lado, Abzalov (2011b). Em (b), um quarteador de cruzeta divide duas porções de maneira que a granulometria média de uma fração é consideravelmente maior que a outra. Em ambos os casos, os resultados das análises serão portadores de vieses.

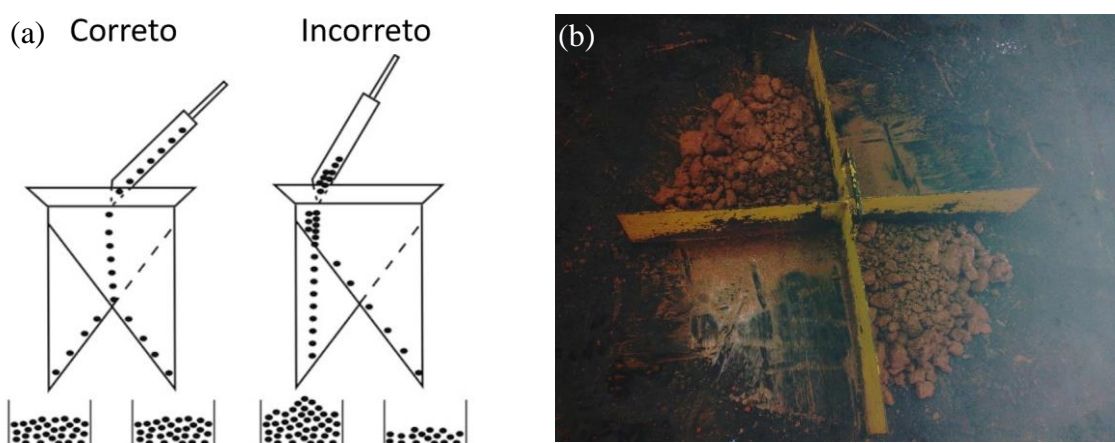


Figura 16: (a) Alimentação correta e incorreta em um quarteador tipo rifles, (b) Divisão enviesada com quarteador de cruzeta.

Fonte (a): modificada, Abzalov (2011b).

Fonte (b): Souza; Costa; Souza (2021).

2.2.2. Teoria da Amostragem Quantitativa

Como exposto anteriormente, materiais naturais apresentam uma heterogeneidade intrínseca no lote. Mesmo minimizando ao máximo erros ligados a amostragem incorreta, ainda existirão os erros relacionados com a variabilidade do material e do processo. Assim, Gy (2004a) descreve três erros que constituem os erros de amostragem correta, Erro de Flutuação de Heterogeneidade (HFE), Erro de Segregação e Agrupamento (GSE) e Erro Fundamental (FSE).

O mesmo autor se refere a dois modelos distintos, Equação 25 para o CSE de modelos Zero-dimensionais e Equação 26 para modelos unidimensionais, Gy (2004c).

$$\mathbf{CSE = FSE + GSE} \qquad \mathbf{Equação 25}$$

$$\mathbf{CSE = HFE + (FSE + GSE)} \qquad \mathbf{Equação 26}$$

2.2.2.1. Erro de Flutuação de Heterogeneidade ou de Seleção Pontual (HFE)

Primeiramente, para compreendermos o HFE, é necessária a definição de dois modelos matemáticos distintos, populações unidimensionais e zero-dimensionais. Para Gy (2004b), uma população de unidades não ordenadas é descrita como zero-dimensional por convenção. Nesse caso, poderia até haver uma correlação espacial entre unidades, mas o modelo não a leva em conta, como sacos com pó de perfuratriz que chegam ao laboratório de análise.

Ainda, conforme o mesmo autor, o modelo de uma dimensão (Gy, 2004c) pode ser definido como uma série de unidades ordenadas, tal qual uma massa de minério em um transportador de correia fluindo no tempo ou corpos unidimensionais extensos.

Os lotes unidimensionais podem sofrer flutuações geradas por variabilidade associada às operações da mina e da planta. Dessa forma, para Pitard (1993), os erros decorrentes dessas flutuações são de longo alcance, têm origem em fenômenos do ambiente e são divididos em duas classes: erros não periódicos, entendidos como tendências e deslocamentos de qualidade conforme Gy (2004c), e erros periódicos, oriundos de fenômenos cíclicos.

Notando o exemplo da Figura 17, um amostrador automático retira amostras de minério cronologicamente de um transportador de correia. Entretanto, o fluxo mássico sofre de flutuações periódicas de teor, o que privilegia a tomada de alíquotas apenas com teores altos. Essa configuração caracteriza a quebra da Lei da Equiprobabilidade na amostragem. Existe um viés resultado do erro de seleção pontual cíclico, ou seja, o resultado da média de teores das amostras é significativamente maior que a média dos teores do lote inteiro e, portanto, tratam-se de amostras não representativas.

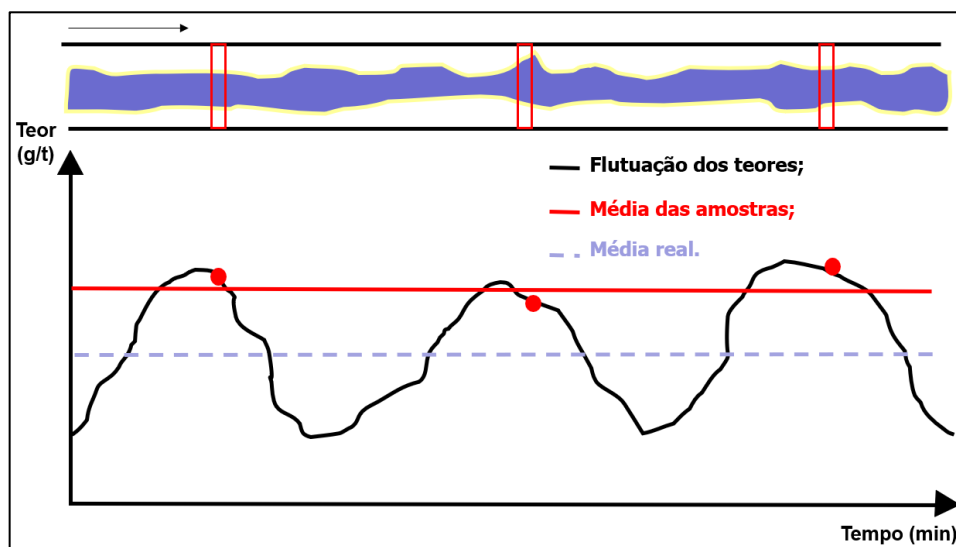


Figura 17: Erro de seleção pontual periódico em transportador de correia.

Fonte: modificada, Souza; Costa; Souza (2021).

De modo semelhante, se a massa de minério transportada pela correia varia mais que 10% relativo com um intervalo de confiança de 68%, conforme Pitard (1993), a tomada dos incrementos pode tornar-se não representativa e geradora de erro. Vieses causados pelas variações de quantidade de material num processo de seleção de incremento são chamados de Erros de Ponderação do Incremento (IWE), conforme Goes & Possa & Luz (1991).

2.2.2.2. Erro de Segregação e Agrupamento (GSE);

Essencialmente, o GSE é uma consequência da heterogeneidade distribucional em si. Para que o GSE fosse zero, Gy (1998) estabelece a condição de que todos elementos

selecionados devem ser independentes um do outro. Entretanto, estamos em um campo gravitacional que age sobre o lote e as partículas não são homogêneas em formato e constituição, dessa forma, essa condição só é cumprida teoricamente.

Gy (1998) completa, na prática, são extraídos incrementos compostos por vizinhos, e são esses grupos que possuem probabilidade de serem selecionados, ou seja, a seleção está correta, mas a condição de independência não é cumprida. Além disso, esse erro é formado por dois fatores, um de segregação, uma medida dos rearranjos espaciais, e outro de agrupamento, que é uma medida da seletividade aleatória. O GSE é proporcional ao tamanho do incremento e a heterogeneidade constitucional (DHL).

Exemplificando este erro, é possível observar a situação de coleta de incrementos em uma pilha de pó de perfuração. Por mais que a coleta seja realizada de forma a reduzir a zero os erros de amostragem incorreta, o material apresentará uma segregação natural e um agrupamento preferencial das partículas, inerentes à constituição e a própria operação da perfuração. A estratigrafia e a densidade dos diferentes minerais podem resultar em segregação e agrupamento de grãos, esquematizados na Figura 18, e conseqüentemente em vieses nos resultados de teores das amostras.

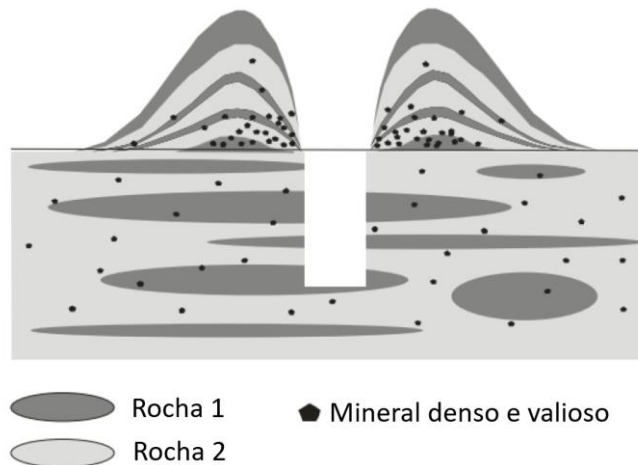


Figura 18: Pilha de pó de perfuração com grãos segregados e agrupados.

Fonte: modificada, Abzalov (2011b).

2.2.2.3. Erro Fundamental de Amostragem (FSE);

De forma simplificada, o erro fundamental determinado por Gy (1998) é um erro que sempre existirá nos resultados da amostragem e é consequência da heterogeneidade do material. Para calcularmos este erro com base no CH_L , tomamos a Heterogeneidade Intrínseca do Lote (IH_L) (Equação 27) como premissa. Ela se trata de uma aproximação prática, pois h_i é raramente estimada devido a sua complexidade, ficando reservada na forma $\sigma^2(h_i)$ como um caráter mais teórico.

$$IH_L = \sum_i \frac{(a_i - a_L)^2 M_i^2}{a_L^2 M_L} \quad \text{Equação 27}$$

Em um lote Zero-dimensional, a Heterogeneidade Invariante foi definida por Gy (1998) como a multiplicação entre a constante de amostragem (C) e o *top size* ao cubo em centímetros (d_N^3) (Equação 28). O *top size* é definido como o diâmetro de malha de peneira em que 95% da massa total é passante, já a Constante de Amostragem é uma multiplicação de outros quatro fatores.

$$IH_L = C * d_n^3 = f * g * c * l * d_n^3 \quad \text{Equação 28}$$

- Fator de Forma (f): adimensional, 0,1 (para partículas lamelares) e 1 (para partículas cúbicas);
- Fator Granulométrico (g): adimensional, entre 0,1 e 1 (quando todas as partículas possuem o mesmo tamanho), 0,25 na maioria dos casos;
- Fator Mineralógico (c): na unidade de g/cm^3 , calculado em função do teor médio do mineral de interesse (a_L) e das densidades específicas no lote do mineral de interesse (ρ_M) e da ganga (ρ_G) segundo Equação 29;

$$c = \left(\frac{1 - a_L}{a_L}\right) \left((1 - a_L) \rho_M + (a_L * \rho_G) \right) \quad \text{Equação 29}$$

- Fator de Liberação (l): adimensional, calculado em função do tamanho de liberação do mineral de interesse (d_l) e do *top size* (d_N) conforme Equação 30.

$$l = \left(\frac{d_l}{d_N}\right)^{0,5} \quad \text{Equação 30}$$

Para evitar o uso de constantes tabeladas, é possível realizar calibração da fórmula com teste de heterogeneidade, François-Bongarçon (2002). Na Equação 31, obtemos o cálculo da Variância do erro fundamental, $\sigma^2(FSE)$, de acordo com Gy (1998), em termos da redução de massa ocorrida entre o lote e amostra. Ao trabalharmos os erros a nível de variância, somos beneficiados por ser uma grandeza aditiva, ou seja, a equação pode ser utilizada na soma de erros das diferentes etapas de um protocolo amostral, ou do GEE, e transformadas em desvios padrões posteriormente.

$$S^2(FSE) = \left(\frac{1}{M_S} - \frac{1}{M_L}\right) H I_L \quad \text{Equação 31}$$

Finalmente, o Erro Total de Amostragem (TSE) é resultado da soma da Equação 32.

$$TSE = ISE + CSE = (IDE + IEE + IPE) + (HFE + (FSE + GSE)) \quad \text{Equação 32}$$

2.2.3. Protocolo amostral

O protocolo amostral, ou de preparação das amostras, pode ser entendido como uma série de etapas, geralmente associadas ao ambiente de laboratório, que objetivam a redução de massa da amostra de forma que a alíquota analisada quimicamente seja representativa do lote. Cada minério necessita de um protocolo amostral específico de maneira a minimizar o erro fundamental.

Um protocolo amostral adequado é aquele que o Erro Total de Amostragem (TSE) está dentro de um intervalo aceitável pelas partes interessadas. Na Figura 19, observamos o exemplo de um típico protocolo de amostragem para material proveniente do ROM (*Run of Mine*). Ao ser coletada na pilha de desmonte, a amostra passa por sucessivas etapas de quarteamento e cominuição (traçado em preto), até atingir massa e granulometria adequadas para a análise química laboratorial.

As etapas de homogeneização e cominuição tem o objetivo de diminuir os efeitos da Heterogeneidade Distribucional do Lote e da Heterogeneidade Constitucional do Lote respectivamente. Entretanto, são necessários mecanismos para o monitoramento e controle da confiabilidade dos dados, sua precisão. Assim, é adotado o sistema de garantia de qualidade (Quality Assurance – QA) e de procedimentos de controle de qualidade (Quality Control – QC), chamado QAQC, Chierigati (2007).

A precisão indica a reprodutibilidade dos resultados analíticos e, na indústria, em muitos casos, o programa QAQC é aplicado por meio da análise de duplicatas (Abzalov, 2008; 2011a) ou quadruplicatas de fim de protocolo, esquematizadas na Figura 19. O erro de precisão é matematicamente deduzido a partir das diferenças entre os pares de amostras (originais mais duplicatas).

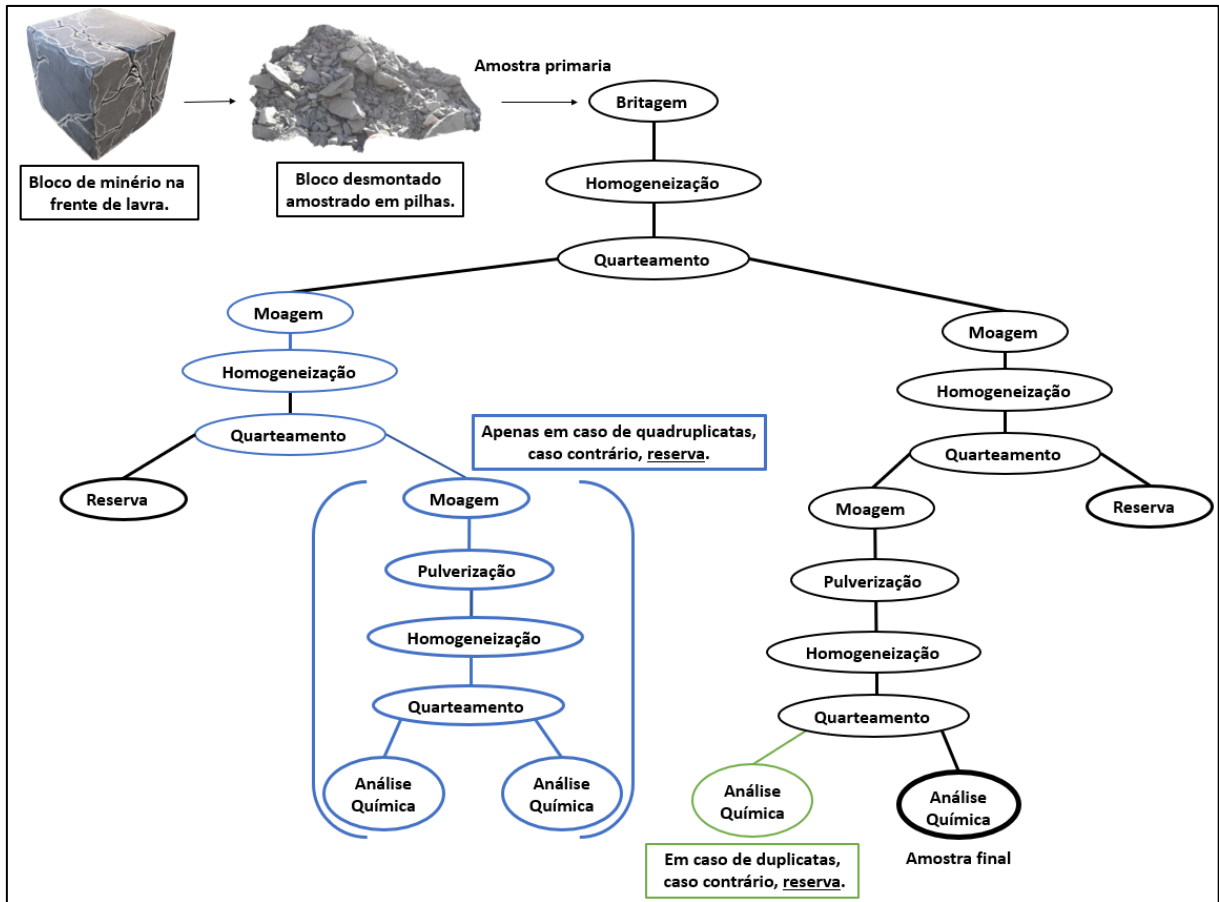


Figura 19: Exemplo de protocolo amostral com hipóteses de análise de duplicatas ou quadruplicatas.

Fonte: o Autor (2022).

Entretanto, mesmo os atuais procedimentos QAQC adotados na indústria não são capazes de garantir a representatividade do lote ou a minimização máxima dos erros de protocolo. Tomando o exemplo da Figura 19, ainda que uma comparação entre os resultados das análises químicas de teores das quatro amostras pertencentes ao conjunto de quadruplicatas demonstrem uma variância pequena, se a tomada de amostra foi realizada em uma pilha de desmonte, o incremento já possui um erro associado inerente. Ou seja, a tomada da amostra em pilha de minério fere a Lei da Equiprobabilidade, porções internas da pilha não são acessíveis, gerando Erros de Delimitação (IDE) e de Extração (IEE), que não podem ser quantificados e acompanharão a amostra inevitavelmente.

Transitando entre contextos de tomadas de amostras minerais, a Figura 20 demonstra um amostrador automático cortando o fluxo da polpa de rejeito de uma usina de beneficiamento de minério de ferro. É notória a falta de limpeza do equipamento entre as tomadas de

incrementos, essa contaminação causa viés na amostra através do Erro de Preparação dos Incrementos (IPE). Novamente, mesmo um protocolo de preparação das amostras otimizado, que não gere diferenças de teor entre duplicata e amostra final no QAQC habitual, já está com o IPE contido no resultado final, há viés.



Figura 20: Amostrador de rejeito de minério de ferro.

Fonte: Souza; Costa; Souza (2021).

Além dos erros citados e exemplificados até aqui, ainda há um outro causador de vieses que o programa QAQC baseado em análises de duplicatas finais não é capaz de considerar. Esse erro tem estreita relação com as reduções de massas do incremento frente a massa do lote. Segundo Armstrong (1950), por exemplo, ao considerarmos o valor do teor de uma amostra de testemunho para um bloco inteiro de minério a ser lavrado, estamos extrapolando a média de teor de um número limitado de incrementos para toda uma população, maior e mais heterogênea. Isto é, estamos cometendo um erro de precisão chamado Erro de Extensão (EE), corresponde ao procedimento de estender os teores de uma amostra a uma zona de influência (Tomazi & Souza, 2021). Para que seja possível conceituar e definir matematicamente o EE, são necessárias algumas concepções de Geoestatística.

2.3.GEOESTATÍSTICA

Para Koppe et al. (2006), a Geoestatística compreende o conjunto de métodos que tem por objetivo a estimativa de valores de um atributo, que são correlacionados no tempo. A coleta de dados amostrais tem um elevado custo e pode sofrer com dificuldades de acesso, assim, a estimativa de valores é necessária em locais não amostrados.

Em um breve histórico, a criação e difusão da Geoestatística é atribuída a quatro pesquisadores, Georges Matheron, Daniel Krige, André Journel e Alain Marechal. Matheron, formado pela École Normale Supérieure des Mines de Paris, criou, em 1968, o Centre de Morphologie Mathématique, que adquiriu status de importante centro de estudo, difusão e formação de pesquisadores de Morfologia Matemática e Geoestatística. André Journel e Michael Davis, ex-alunos de Matheron, foram os responsáveis por sua difusão na América do Norte, publicando as famosas obras *Geostatistical Ore Reserve Estimation*, de 1977, e *Mining Geostatistics*, em 1978.

É importante ressaltar que a Geoestatística não é uma ciência exata, não fornece exatamente a quantidade de minério contido em determinado local. Sempre existirá um erro associado, fruto do método de interpolação utilizado. Esse erro deve ser minimizado para que os resultados se aproximem o máximo possível da realidade. A Geoestatística jamais substitui dados adicionais, nem o bom entendimento e correto julgamento do problema físico em questão.

2.3.1. Variografia

A variografia é a principal ferramenta utilizada na geociência para medir a continuidade espacial. Por meio do variograma, é possível determinar uma equação matemática que descreve bem o comportamento do fenômeno geológico de ocorrência, mostrando em quais direções e a que distância existe maior correlação entre as amostras. O variograma mede o grau de diferença entre um valor de um local não amostrado e uma amostra próxima deste local. Esta função também é a base da krigagem, o algoritmo mais comum para definir (estimar) teores (Goovaerts, 1997). Diferente de outras áreas da Estatística aplicada, as quais usam normalmente a covariância e as funções de correlação para caracterizar continuidade, a Geoestatística utiliza o momento de inércia. A função variograma pode ser definida pela Equação 33, segundo Journel e Huijbregts (1978, apud Yamamoto & Landim, 2013), como o quadrado da esperança matemática entre pares de pontos separados por uma distância h .

$$2\gamma(\mathbf{h}) = E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\} \quad \text{Equação 33}$$

Segundo o mesmo autor, com o intuito de facilitar o entendimento do cálculo, escreve-se a esperança matemática como um somatório e com a incógnita isolada, temos a equação da metade do variograma, chamada semivariograma (Equação 34), isto é, média da metade das diferenças ao quadrado dos pontos separados por uma distância h .

$$\gamma(\mathbf{h}) = \frac{1}{2 \cdot n} \sum_{i=1}^n [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2 \quad \text{Equação 34}$$

Onde:

$\gamma(\mathbf{h})$: semivariograma;

n : número de pares de amostras;

$Z(\mathbf{x}_i)$: amostras localizadas espacialmente em determinado local;

$Z(\mathbf{x}_i + \mathbf{h})$: amostras separadas de uma distância h de $Z(\mathbf{x}_i)$.

O Semivariograma Relativo $\gamma_R(\mathbf{h})$ é definido por Abzalov (2008) como a divisão das diferenças entre os pares de dados pela média aritmética global da variável estudada ao quadrado (Equação 35). Esse conceito tem grande importância no estudo das variâncias das duplicatas do protocolo amostral.

$$\gamma_R(\mathbf{h}) = \frac{1}{2 \cdot n} \sum_{i=1}^n \frac{[Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2}{m^2} \quad \text{Equação 35}$$

Na literatura, é comum chamar o semivariograma apenas de variograma, o que será uma prática adotada daqui para frente nesse trabalho. A apresentação do semivariograma é feita

através de um gráfico, e a Equação 34 gera cada ponto desse gráfico para todos valores de h . É esperado que, com o aumento da distância, a variabilidade aumente até o momento em que o aumento da distância não irá mais influenciar na diferença quadrática dos pares. Ou seja, quanto maior a dispersão, maior o momento de inércia e menor a correlação (Yamamoto & Landim, 2013).

Em Estatística, a covariância é uma medida da relação mútua entre duas variáveis aleatórias distintas, mas na Geoestatística, de acordo com Yamamoto & Landim (2013), a covariância mede a relação entre valores da mesma variável, obtidos em pontos separados por uma distância h , conforme uma determinada direção. Quando alteramos a direção e a covariância se altera, podemos estar diante de um evento espacial anisotrópico, caso a covariância permaneça a mesma em qualquer direção, é um fenômeno isotrópico. Assim, para Goovaerts (1997), a covariância $C(h)$ de uma variável regionalizada x para pontos separados por uma distância h pode ser calculada com a Equação 36:

$$C(h) = E\{[Z(x + h) - m][Z(x) - m]\} \quad \text{Equação 36}$$

Onde:

h : representa um vetor entre x_1 e x_2 no espaço tridimensional;

Conforme Yamamoto & Landim (2013), se o vetor h apresentar-se infinitamente pequeno, a variância será mínima e a covariância, máxima. Haverá um valor Δh para o qual as duas podem apresentar valores aproximadamente iguais, porém, à medida que Δh aumenta, a covariância diminui enquanto a variância aumenta, pois ocorre progressivamente maior independência entre os pontos a distâncias cada vez maiores. Esse comportamento é possível observar no Gráfico 8 de variograma ou correlograma, gráfico que possui a correlação espacial no eixo vertical e distância espacial no eixo horizontal.

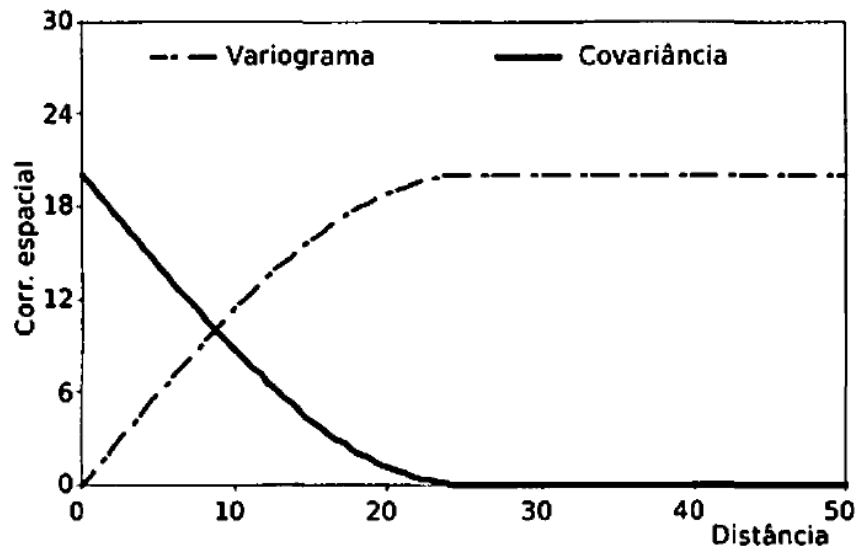


Gráfico 8: Gráfico semivariograma e covariância.

Fonte: Yamamoto & Landim (2013).

O gráfico do variograma possui alguns elementos importantes que são determinantes em sua interpretação (Gráfico 9):

- i. *Sill* (patamar): é o platô atingido pelo variograma conforme aumenta a distância h . Seu valor se aproxima da variância total dos valores amostrais (variância *a priori* dos dados);
- ii. *Range* (alcance): distância na qual o variograma atinge o patamar;
- iii. *Lag*: intervalos entre as tomadas de amostras;
- iv. *Nugget Effect* (efeito pepita), $V(0)$, ou variância *a priori*: é o salto vertical na distância zero, origem, para o valor do variograma. Embora teoricamente $\gamma(0)=0$, erros de amostragem somados a erros de medição acabam gerando essa descontinuidade na origem do variograma.

Até esse momento, tratou-se da variografia em termos de variância no espaço, entretanto, para este estudo, é relevante a compreensão da cronovariografia, variografia em termos de tempo. Nessa abordagem, a variância diz respeito aos intervalos de tomadas de amostras de minério em transportador de correia ao longo do tempo. Sempre que possível, será traçado esse paralelo entre a variância espacial e temporal.

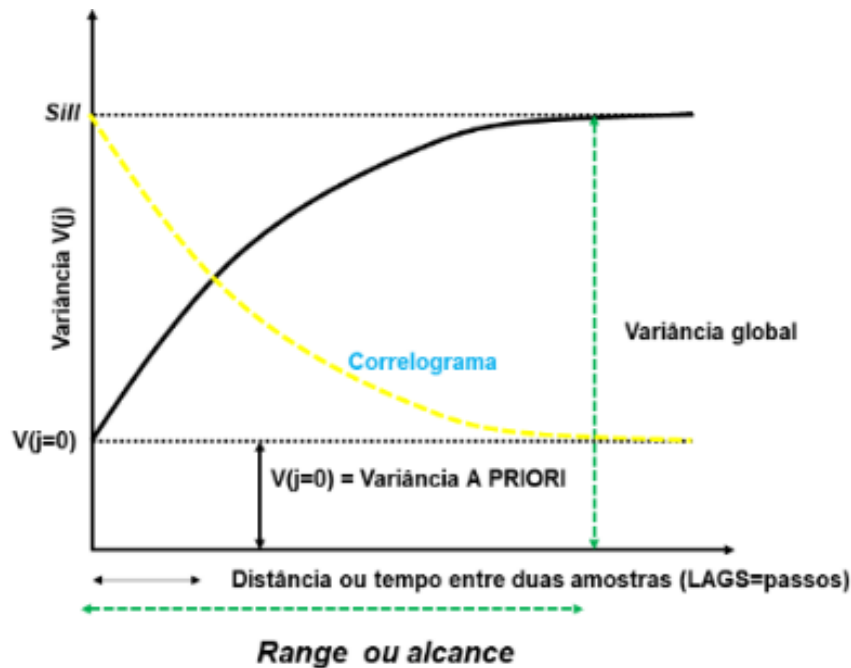


Gráfico 9: Exemplo de cronovariograma com elementos destacados.

Fonte: Tomazi (2021).

Mais importante que medir a continuidade do fenômeno, é poder determinar a direção na qual este fenômeno é mais contínuo. No caso da variância no tempo, possuímos apenas a direção espacial do sentido do fluxo da correia. Para tanto, existem alguns parâmetros de busca que são utilizados quando os pares de amostras estão sendo procurados, de forma a restringir esta busca a determinados setores. Deste modo, todos os parâmetros restritivos presentes na Tabela 4 serão considerados no cálculo do variograma e cronovariograma. O Gráfico 10 representa uma busca variográfica espacial com os seus elementos.

Tabela 4: Relação entre parâmetros de um variograma e cronovariograma.

Parâmetro	Variograma espacial	Variograma temporal
<i>Lag</i>	Distância do vetor com que se deve procurar pares de amostras;	Número de amostras tomadas;
Separação do <i>lag</i>	Distância entre as amostras;	Tempo entre as coletas;
Tolerância do <i>lag</i>	Delimita um intervalo em torno do <i>lag</i> e, em geral, equivale à metade do próprio <i>lag</i> ;	
Azimute	Direção da busca;	Direção da busca no sentido da correia;

<i>Dip</i>	Mergulho da busca;	Por tratar-se de apenas uma dimensão, é considerado zero;
Tolerância do Azimute ou Tolerância Angular;	Ângulo definido para evitar a busca de pares a distâncias muito curtas que não estejam na direção requerida;	Igual ao próprio Azimute;
<i>Bandwidth</i>	Largura da busca.	Largura da correia.

Fonte: o Autor (2022).

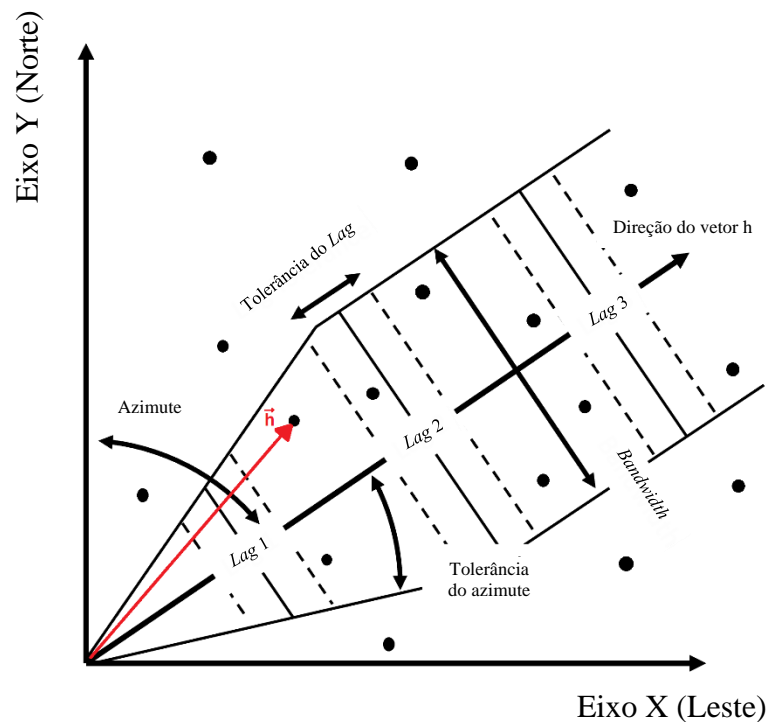


Gráfico 10: Cone de busca do variograma com principais elementos.

Fonte: modificado, *Pythonic Excursions* (2020).

Uma vez definidos todos os parâmetros de busca, o próximo passo é determinar um modelo variográfico que represente a variabilidade do fenômeno de interesse. O objetivo do modelamento é fornecer valores para qualquer *lag* possível que atenuem a oscilação dos dados através de uma função contínua. Mas, segundo Goovearts (1997), é importante notar que nem todas as funções contínuas podem ser admitidas no variograma, somente aquelas que sejam sempre positivas e crescentes. Entre os possíveis tipos estão: esférica, exponencial e gaussiana, conforme o Gráfico 11. O variograma pode conter mais de um tipo.

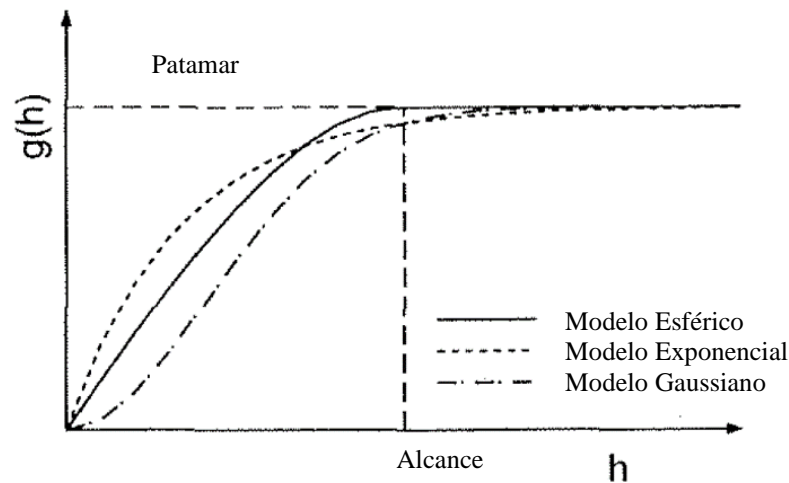


Gráfico 11: Funções que podem constituir o variograma.

Fonte: modificado, Goovaerts (1997).

O variograma, de forma geral, é sensível a valores extremos e agrupamentos de dados combinados com o efeito proporcional. O efeito proporcional é a variância local dos dados relativa à sua média, podendo possuir a assimetria positiva (média elevada, variância elevada) ou negativa (média baixa, variância elevada).

2.3.1.1. Análise de Duplicatas e Efeito Pepita

Como discutido anteriormente, o programa QAQC, usualmente, leva em conta as diferenças entre os pares de amostras finais, duplicatas (“a” e “b”). Quando os dados apresentam uma distribuição fortemente assimétrica e contém *outliers*, Goovaerts (1997) sugere utilizar o Variograma Relativo dos Pares $\gamma_{PWR}(h)$ (*Pair-Wise Relative variogram - PWR*) para quantificar a continuidade. Ele é calculado dividindo o semivariograma pela média aritmética ao quadrado dos pares correspondentes dos dados (Equação 37).

$$\gamma_{PWR}(h) = \frac{1}{2n} \sum_{i=1}^n \frac{[Z(x_i) - Z(x_i + h)]^2}{\left[\frac{Z(x_i) + Z(x_i + h)}{2}\right]^2} = \frac{1}{2n} \sum_{i=1}^n \frac{[Z(x_i) - Z(x_i + h)]^2}{[Z(x_i) + Z(x_i + h)]^2} \quad \text{Equação 37}$$

Assim, o erro de precisão, a partir da diferença entre os pares de amostras, pode ser calculado através do variograma relativo de pares (PWR) aplicado a pares de duplicatas ($h=0$), definido por Abzalov (2008) na Equação 38. Sua vantagem é poder comparar erros de amostragem e fazer somas. Já a Equação 39 define o coeficiente de variação (CV%) e foi proposta como uma medida universal da precisão relativa do erro em aplicações em mineração, Stanley & Lawie (2007 apud Abzalov, 2008).

$$\sigma^2_{RSV} = \frac{2}{n} \sum_{i=1}^n \frac{[Z(x_i) - Z(x_i + 0)]^2}{[Z(x_i) + Z(x_i + 0)]^2} = \frac{2}{N} \sum_{i=1}^N \left(\frac{(a_i - b_i)^2}{(a_i + b_i)^2} \right) \quad \text{Equação 38}$$

$$\begin{aligned} CV\% &= 100\% \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{\sigma_i^2}{m_i^2} \right)} \\ &= 100\% \sqrt{\frac{2}{N} \sum_{i=1}^N \left(\frac{(a_i - b_i)^2}{(a_i + b_i)^2} \right)} \end{aligned} \quad \text{Equação 39}$$

O modelo PWR é capaz de inferir o efeito pepita, o qual representa a combinação dos erros de amostragem e descontinuidade geológica a curtas distâncias ($h \rightarrow 0$) (Abzalov, 2008). Dessa forma, quando subtraímos o valor da variância do erro fundamental da amostragem σ^2 (FSE) do efeito pepita, obtemos puramente a contribuição da heterogeneidade do minério originada de fatores geológicos. Esse comportamento pode ser observado na representação da Figura 21.

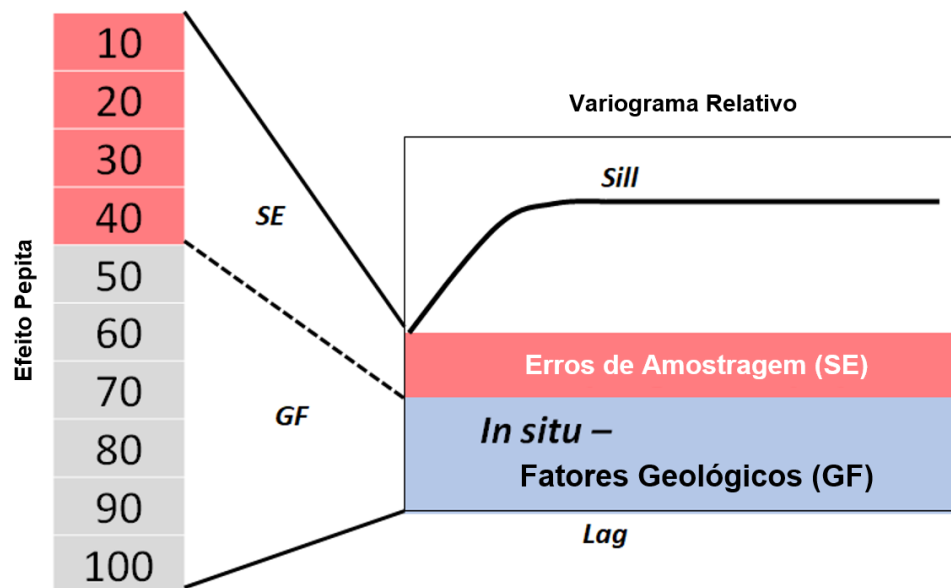


Figura 21: Exemplo de contribuição dos erros de amostragem no aumento do efeito pepita do variograma.

Fonte: modificada, Abzalov (2008).

De posse da dimensão da contribuição dos erros de amostragem no aumento do efeito pepita, é possível modificar os protocolos de amostragem com o objetivo de minimizar a variância dos erros (Abzalov, 2008). Esse procedimento permite investigar fontes de erros, melhorar a confiabilidade das estimativas de recursos e otimizar o beneficiamento. Ainda assim, quando adotadas melhorias no protocolo de amostragem, dificilmente é levado em consideração um importante contribuinte para o inchaço da variância em pequenas distâncias, o Erro de Extensão (EE).

2.3.2. Erro de Extensão (EE)

Esse erro é cometido com respeito a toda massa do lote que não é coletada durante a amostragem. Seu conhecimento é essencial nos processos de reconciliação proativa (Chierigati, 2008) e ajuda a desenhar os circuitos e equipamentos de uma usina de tratamento de minério (Pitard, 1993). Apesar deste poder ser o maior erro de amostragem, quase sempre, ele é ignorado. Ainda, quando os erros de protocolo de preparação das amostras são determinados, junto com o erro analítico, estas incertezas não retroalimentam os modelos de reconciliação

(Abzalov, 2011a, 2016; Chierregati et al. 2008; Chierregati & Pitard, 2009; Marques & Costa, 2013, 2014a, 2014b).

Dessa maneira, a variância de extensão refere-se ao erro de precisão cometido ao diminuir o número de incrementos e extrapolar a média, variância ou outra medida dos teores ou de outro parâmetro dos incrementos analisados para toda população. Exemplificando, quando extrapolamos o valor médio dos teores dos incrementos coletados automaticamente em intervalos horários sobre um transportador de correia para todo o fluxo mássico que passou por aquela correia no turno, dia ou mês e assim por diante, estamos cometendo um erro de extensão. Quanto maior o intervalo de tempo entre coletas, maior é este erro.

Estes intervalos de tempo são chamados *lags* ou passos (h). “Seria esperado” que em *lag* igual a 0 (zero), a variância fosse zero, porém, como já visto no item sobre o efeito pepita, isto não ocorre. Este valor de variância em *lag* zero é maior quanto mais heterogêneo for o lote e é “inflado” devido aos erros de amostragem (Pitard, 1993). Para calcular a variância para um conjunto de amostras de *lag* igual a *j* minutos, adotamos a Equação 40, o teor da amostra “*a_i*” está “separado *j* minutos” da amostra de teor “*a_{i+j}*”, sendo *n* o número de amostras e *n-j* igual ao número de *lags* ou passos.

$$\sigma^2(j) = \frac{1}{2(n-j)} \sum_{i=1}^{n-j} (a_{i+j} - a_i)^2 \quad \text{Equação 40}$$

A variância de extensão pode ser considerada como um erro de longo alcance ou grande escala (Pitard, 1993, 1994). Ao assumir que a amostra representa uma faixa maior da correia, que é a zona de influência (Figura 22), comete-se um erro quantificável através da Cronoestatística, mais especificamente a variografia. Este erro é denominado por alguns autores como erro de interpolação (Pitard, 1993, 1994; Mucha & Szuwarzyński, 2004; Esbensen & Engström 2017).

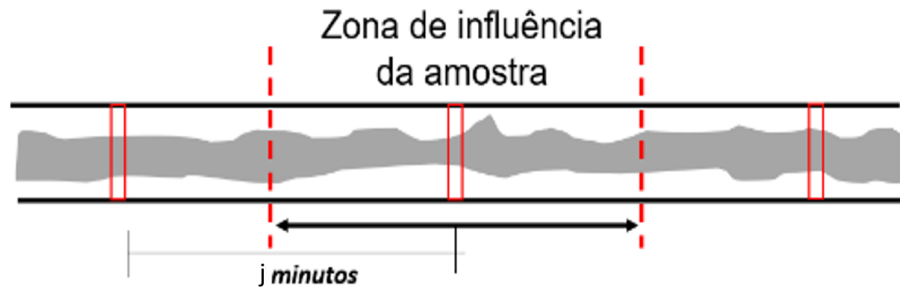


Figura 22: Zona de influência da amostra tomada sobre transportador de correia.

Fonte: o Autor (2021).

A variabilidade da zona de influência (“variabilidade real”) tende a ser maior do que quaisquer intervalos menores que possam ser considerados. No esquema do transportador de correia da Figura 23, trechos em amarelo, verde, vermelho, azul e cinza, representam faixas de minério com teores muito diferentes entre si. Considerando como j minutos o tempo entre coletas de uma amostragem, os incrementos serão tomados nos três locais representados pelos retângulos vermelhos mais curtos e estreitos. Dessa forma, os resultados das análises de teores das amostras coletadas demonstrarão uma imprecisão “pequena” (apenas teores cinza e amarelo coletados), podendo parecer preciso o suficiente e acurado o suficiente para um objetivo específico. Entretanto, conforme a mesma figura, a amostragem não foi representativa do lote real. (Matheron, 1971, 1984; Journel & Huijbregts, 1978; Mucha & Szuwarzyński, 2004; Marques & Costa, 2013, 2014a, 2014b; Wackernagel, 1998).

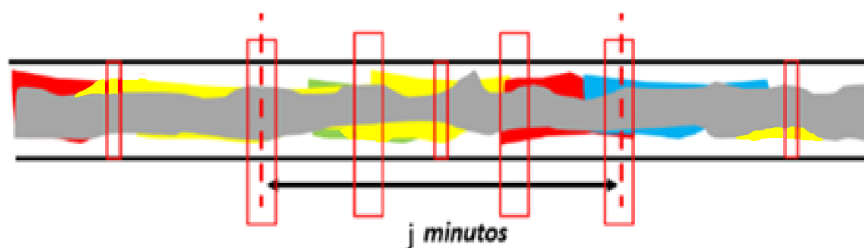


Figura 23: Exemplo de erros de interpolação.

Fonte: o Autor (2022).

A Equação 41 é utilizada para o cálculo da variância de extensão segundo Wackernagel (1998). Nela, a variância de extensão é $\sigma_{ext}^2 \left(\frac{\cdot}{T} \right)$; $\underline{\gamma} \left(\frac{\cdot}{t} \right)$ é a variância de dispersão média com

respeito ao intervalo de tempo menor “t” e $\underline{\gamma}\left(\frac{\cdot}{t}\right)$ é a variância média de dispersão com respeito ao intervalo de tempo maior.

$$\sigma_{ext}^2\left(\frac{\cdot}{T}\right) = 2 \cdot \underline{\gamma}\left(\frac{\cdot}{t}\right) - \underline{\gamma}\left(\frac{\cdot}{T}\right) \quad \text{Equação 41}$$

Ainda, “T” é o intervalo de tempo que se quer estimar a variância de extensão, um intervalo qualquer maior que o intervalo “t” em que se deixará de coletar amostras e por isso estaremos cometendo um erro. O intervalo “T” de tempo é o dobro do intervalo de tempo anterior “t”. O menor intervalo de tempo em que é possível coletar amostras é representado por um “.” (“ponto”) nesta equação. Sua dedução será explanada no capítulo da metodologia.

3. METODOLOGIA

3.1.DEDUÇÃO DA FÓRMULA DA VARIÂNCIA DE EXTENSÃO

Conforme o conceito geral explicado anteriormente, o primeiro passo da metodologia será a dedução da equação do Erro de Extensão a partir da relação de Krige (Wackernagel, 1998). Conforme a Tabela 5, se fossem coletados 12 incrementos ao longo transportador de correia de uma usina de beneficiamento de minério de ferro em um intervalo de 4 em 4 minutos, após 48 minutos, seriam obtidas 12 amostras e suas respectivas análises de teores posteriormente, tal como Krige pensou quando estava estimando os teores para uma bancada na frente de lavra.

Tabela 5: Amostras de transportadora de correia.

Incremento											
1	2	3	4	5	6	7	8	9	10	11	12
Tempo (min)											
4	8	12	16	20	24	28	32	36	40	44	48
Teores (%)											
26	27	30	33	30	34	36	32	43	41	42	46

Fonte: o Autor (2022).

Se forem analisadas a média dos teores dessas 12 amostras (Equação 43) e a variância de cada uma delas com respeito a essa média, é obtido o valor chamado de variância de dispersão absoluta (Equação 43). De modo geral, essa variância representaria a precisão do menor intervalo amostral que seria possível de ser aplicado na prática. Neste exemplo, portanto, a variância de dispersão do lote (L) é aquela correspondente à amostragem de 4 em 4 minutos, o qual é o intervalo considerado como mais curto possível de ser executado na prática. A notação dessa variância é VAR(·,L).

$$m = \frac{26 + 27 + 30 + 33 + 30 + 34 + 36 + 32 + 43 + 41 + 42 + 46}{12} = 35 \% \quad \text{Equação 42}$$

$$\sigma_{DISP}^2 = VAR(\cdot, L) = ((26 - 35)^2 + (27 - 35)^2 + (30 - 35)^2 + (33 - 35)^2 + (30 - 35)^2 + (34 - 35)^2 + (36 - 35)^2 + (32 - 35)^2 + (43 - 35)^2 + (41 - 35)^2 + (42 - 35)^2 + (46 - 35)^2) / 12 = 40 \quad \text{Equação 43}$$

Quando as amostras são agrupadas de 8 em 8 minutos, calculando a variância de tal forma a levar em consideração a diferença entre o teor de cada amostra em relação à média de cada um desses blocos, é possível obter a chamada Variância Intrabloco (Equação 44). Entenda-se por bloco aqui o intervalo de 8 em 8 minutos, dobro do tempo do menor intervalo. É, portanto, um novo suporte amostral cujo intervalo ou tamanho do bloco será chamado de “b”.

$$VAR_{INTRA} = ((26 - 26,5)^2 + (27 - 26,5)^2 + (30 - 31,5)^2 + (33 - 31,5)^2 + (30 - 32)^2 + (34 - 32)^2 + (36 - 34)^2 + (32 - 34)^2 + (43 - 42)^2 + (41 - 42)^2 + (42 - 44)^2 + (46 - 44)^2) / 12 = 2,6 \quad \text{Equação 44}$$

Já quando amostras de 8 em 8 minutos são agrupadas calculando a variância de tal forma a levar em consideração a diferença entre o teor de cada amostra em relação à média de todas as amostras, podemos obter a chamada Variância Interbloco (Equação 45).

$$VAR_{INTER} = ((26,5 - 35)^2 + (31,5 - 35)^2 + (32 - 35)^2 + (34 - 35)^2 + (42 - 35)^2 + (44 - 35)^2) / 6 = 37,4 \quad \text{Equação 45}$$

Krige percebeu, portanto, que a soma da variância intrabloco e interbloco é sempre igual a variância de dispersão ($VAR_{INTRA} + VAR_{INTER} = VAR_{DISP}$). É com base nessa relação que é possível realizar a mudança de suporte quando são realizadas as estimativas do depósito em termos espaciais, sendo que, nesse caso, foi adaptado para o exemplo de intervalos temporais, pois é uma amostragem em transportador de correia.

Ao se observarem, de acordo com a notação adotada, os cálculos e relação abaixo (Equação 46), é notório que, quando amostras equivalentes a tomadas em intervalos de 8 minutos são agrupadas, a variância interbloco desses grupos é menor em comparação com a variância de dispersão absoluta. Caso não houvesse a tomada de 4 em 4 minutos, essa variância poderia ser assumida como a lote, ocasionando em um distanciamento maior da representatividade.

$$\begin{aligned} VAR_{DISP}(\cdot, L) &= \text{Var. de 8 em 8 min (intrabloco)} \\ &+ \text{Var. de 8 em 8 min (interbloco)} \\ &= VAR_{INTRA}(\cdot, b) + VAR_{INTER}(b, L) \end{aligned} \quad \text{Equação 46}$$

Onde:

$$VAR_{INTRA}(\cdot, b)=2,6;$$

$$VAR_{INTER}(b, L)=37,4;$$

$$VAR_{DISP}(\cdot, L)=40;$$

Outra vez, para fundamentar a dedução da fórmula da variância de extensão, as amostras são agrupadas em intervalos de 16 em 16 minutos (dobro do intervalo de tempo anterior). Desse fato, surge o valor multiplicador “2” da fórmula da variância de extensão, resultado de uma mudança de suporte (Equação 41). Novamente, neste exemplo, ela ocorre de 4 em 4 minutos, depois de 8 em 8 e, por fim, de 16 em 16 minutos. Em seguida, isso ficará mais claro: observe a Tabela 6 a seguir e os respectivos cálculos das variâncias intra e interblocos, Equação 47 e Equação 48.

Tabela 6: Amostras de transportadora de correia em diferentes intervalos.

Incremento											
1	2	3	4	5	6	7	8	9	10	11	12
Tempo (min)											
4	8	12	16	20	24	28	32	36	40	44	48
Teores amostras de 4 em 4 minutos (g/t)											
26	27	30	33	30	34	36	32	43	41	42	46
Média de teores dos agrupamos amostras de 8 em 8 minutos (g/t)											
26,5		31,5		32		34		42		44	
Média de teores dos agrupamos amostras de 16 em 16 minutos (g/t)											
29				33				43			

Fonte: o Autor (2022).

$$VAR_{INTRA} = ((26 - 29)^2 + (27 - 29)^2 + (30 - 29)^2 + (33 - 29)^2 + (30 - 33)^2 + (34 - 33)^2 + (36 - 33)^2 + (32 - 33)^2 + (43 - 43)^2 + (41 - 43)^2 + (42 - 43)^2 + (46 - 43)^2) / 12 = 5,33 \quad \text{Equação 47}$$

$$VAR_{INTER} = ((29 - 35)^2 + (33 - 35)^2 + (43 - 35)^2) / 3 = 34,67 \quad \text{Equação 48}$$

A relação de adição de Krige para esse novo intervalo amostral chamado de B é mostrada logo abaixo na Equação 49.

$$\begin{aligned}
 VAR_{DISP}(\cdot, L) &= \text{Var. de 16 em 16 min (intrabloco)} \\
 &+ \text{Var. de 16 em 16 min (interbloco)} \\
 &= VAR_{INTRA}(\cdot, \mathbf{b}) + VAR_{INTER}(\mathbf{b}, L)
 \end{aligned}
 \tag{Equação 49}$$

Onde:

$$VAR_{INTRA}(\cdot, B)=5,33;$$

$$VAR_{INTER}(B, L)=34,67;$$

$$VAR_{DISP}(\cdot, L)=40;$$

Conforme os cálculos acima, quando os intervalos de tempo são dobrados, a variância interblocos começa a ficar cada vez menor. De qualquer forma, ambas as variâncias, intra e interblocos, simbolizam alterações de suporte: do “.”, que simboliza o suporte original com o menor intervalo (4 em 4 minutos), para o suporte de 8 minutos e, por fim, para o suporte de 16 minutos. Assim, tem-se que estabelecer a variância de teor do lote com a variância das amostras coletadas em intervalos de tempo cada vez maiores, incorre justamente no Erro de Extensão. Na prática, amostrar em intervalos de tempo cada vez maiores pode resultar em valores de variância do lote menores do que os verdadeiros.

Retomando, para sanar a questão “Por que os cálculos de erro de extensão são realizados dobrando o intervalo de tempo?”: a variância é o somatório das diferenças quadráticas entre o teor de cada amostra e a média de todas as amostras, dividido pelo número total de amostras (Equação 4). Portanto, quando o intervalo de tempo é dobrado, o número de amostras total é dividido por 2, o que justifica a presença do multiplicador “2” na fórmula da variância de extensão (Equação 41).

Para o cálculo da variância de extensão, também é possível adotar a notação da Equação 50. Na fórmula, o termo *gammabar* é o nome do algoritmo do *Software GSLib* que permite calcular a variância média para cada mudança de suporte, fazendo a determinação do variograma médio. Conforme estabelecido, “.” é o intervalo de tempo de 4 minutos (menor intervalo aplicado na prática), “b” é o intervalo de tempo de 8 minutos e “B” é o intervalo de 16 minutos.

$$VAR_{EXT}(\cdot, B) = 2 * \text{gammabar}(\cdot, b) - \text{gammabar}(\cdot, B) \quad \text{Equação 50}$$

Onde:

$VAR_{EXT}(\cdot, B)$: variância de extensão para o intervalo B (que se deseja estimar o EE);

$\text{gammabar}(\cdot, b)$: variância média de dispersão com respeito ao intervalo de tempo b ($b=B/2$);

$\text{gammabar}(\cdot, B)$: variância média de dispersão com respeito ao intervalo de tempo B.

Então, a variância de extensão para “B” é a diferença entre duas vezes a variância média do suporte de menor intervalo praticado, neste exemplo 4 em 4 minutos, com respeito ao intervalo de 8 em 8 minutos, menos a variância média do suporte de 4 em 4 minutos com respeito ao intervalo que se deseja estimar o EE, ou seja, 16 em 16 minutos. Ao final, ainda é possível gerar um gráfico que ilustra o quanto a variância de extensão cresce à medida que o intervalo de tempo de coleta aumenta.

3.2.PROCEDIMENTO PARA CÁLCULO COMPUTACIONAL DO ERRO DE EXTENSÃO

Para que o leitor aprenda em detalhes e possa reproduzir a metodologia, será relatado o passo a passo dos procedimentos computacionais utilizados no cálculo da variância de extensão. Foram utilizados os *softwares* *GSLib* com o algoritmo chamado *Gammabar 2.0*, *AR2GeMS* e, para o gráfico final, *Microsoft Excel 2019*. O objetivo é obter a variância de extensão de teores em intervalos de amostragem de 15, 30, 45, 60, 120 e 180 minutos e seus respectivos erros com intervalo de confiança de 99,7%.

Primeiramente, o banco de dados original tem o formato txt. e é composto por três colunas. A primeira, nomeada de x, diz respeito a 40 registros de tempo em minutos em que amostras de carvão foram coletadas em uma correia transportadora, valores crescentes de 4 em 4 minutos. A segunda coluna, nomeada de y, é composta apenas pelo número 1, pois a correia se trata de um modelo unidimensional. Por fim, a última coluna contém o resultado das análises

químicas dos teores em porcentagem de Sílica nas 40 amostras. Esse conjunto de dados é o resultado de um chamado teste variográfico, que terá uma seção dedicada nesse trabalho.

O banco de dados deve ser carregado no *software AR2GeMS*, para isso, é preciso clicar na aba *Objects*, em seguida na opção *Load Object* (Figura 24). Na janela aberta, o arquivo do banco de dados deve ser selecionado, em seguida, a opção *point set* também, para que as informações sejam carregadas como pontos no espaço (Figura 25). Na próxima janela, o arquivo precisa ser nomeado e as colunas relacionadas como variáveis x, y e z (Figura 26). Os dados carregados tomam o arranjo espacial de uma linha, pois, novamente, trata-se de amostragem em transportador de correia, unidimensional. Na Figura 27, é possível observar esse arranjo e a legenda de teores de sílica.

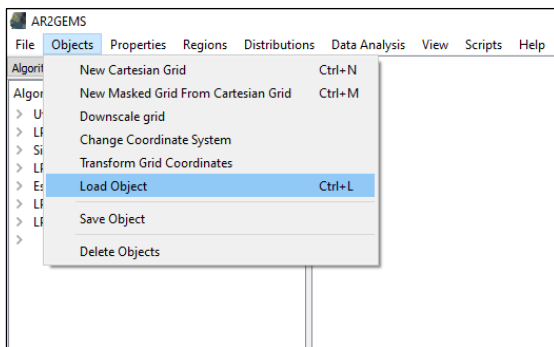


Figura 24: Carregando banco de dados.

Fonte: o Autor (2022).

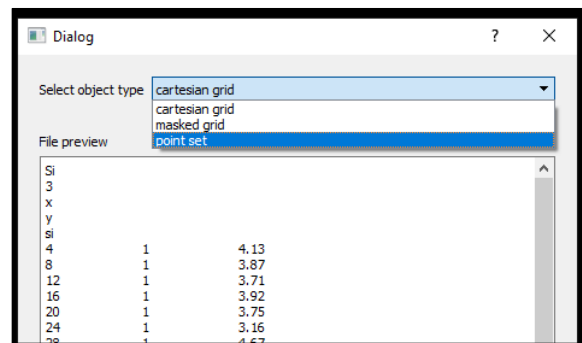


Figura 25: Seleção de tipo de objeto.

Fonte: o Autor (2022).

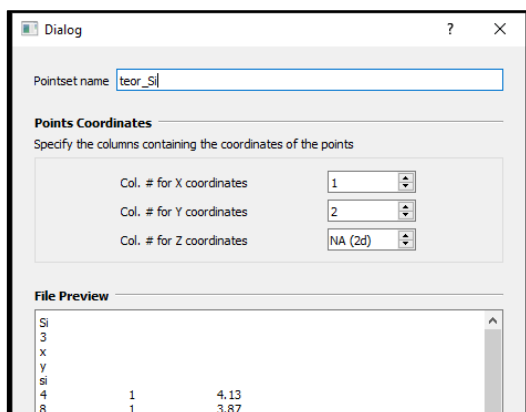


Figura 26: Determinação das variáveis por coluna.

Fonte: o Autor (2022).

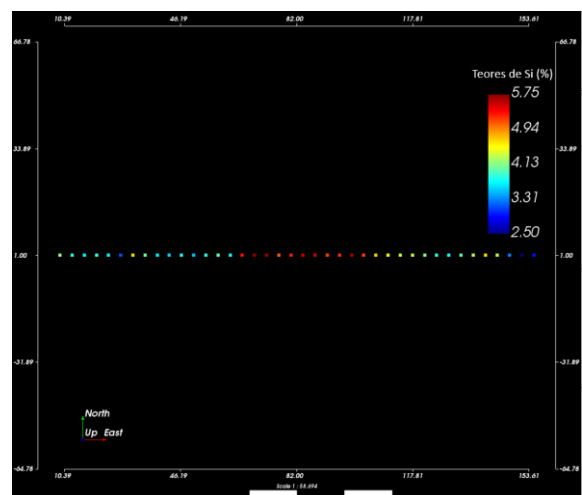


Figura 27: Arranjo dos dados espacialmente.

Fonte: o Autor (2022).

Nesse momento, é importante obter a variância e a média dos dados através do sumário estatístico, os valores, nesse caso, são respectivamente 0,6678 %² e 4,31%. Clicando com o botão direito do mouse nos dados carregados e selecionando a opção *Histogram*, é possível visualizar o histograma e o sumário estatístico do banco de dados (Figura 28).

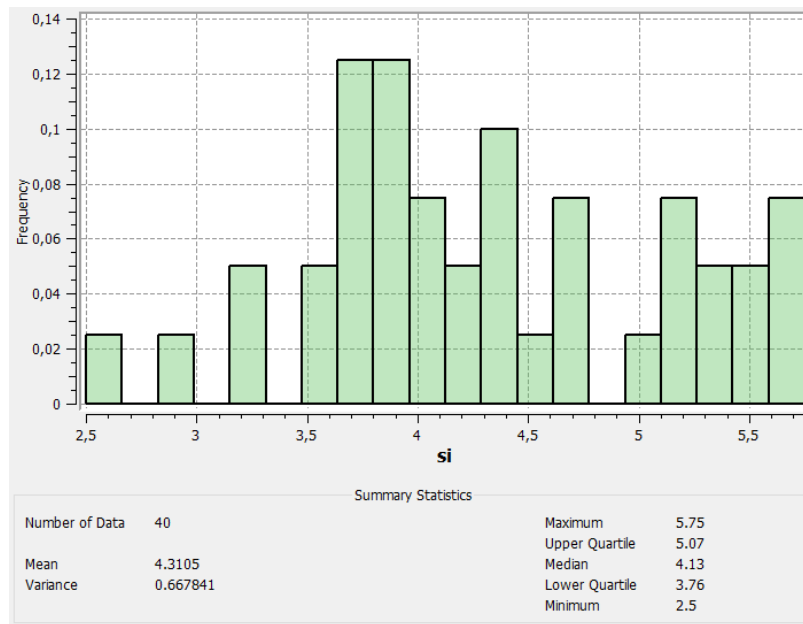


Figura 28: Histograma e sumário estatístico dos dados.

Fonte: o Autor (2022).

Dando início à confecção do variograma, tem de se clicar na aba *Data Analysis* e em seguida em *Variogram* (Figura 29), na janela que foi aberta, deve-se selecionar o arquivo nomeado anteriormente (Figura 30). Na próxima etapa, os parâmetros do variograma são definidos (Figura 31), assim, o número de *lags* corresponde ao número de amostras, 40 nesse caso, a separação do *lag* diz respeito ao menor intervalo de coleta, 4 minutos, e a tolerância do *lag* é a metade desse valor, 2 minutos. Ainda, existe uma única direção de busca com azimute em 90°, devido a direção de fluxo da correia leste-oeste, mergulho de 0°, *bandwidth* de 10 e tolerância de 90°. Embora não modifique os resultados, a tolerância poderia ter um ângulo menor por se tratar de uma busca em linha.

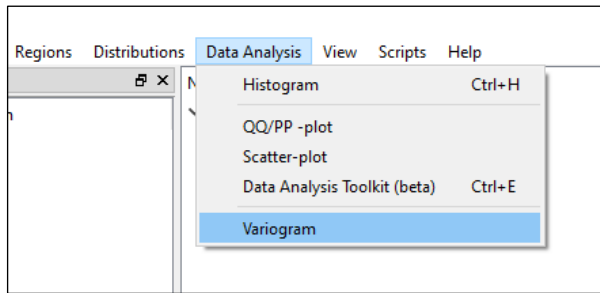


Figura 29: Ferramenta de Variograma.

Fonte: o Autor (2022).

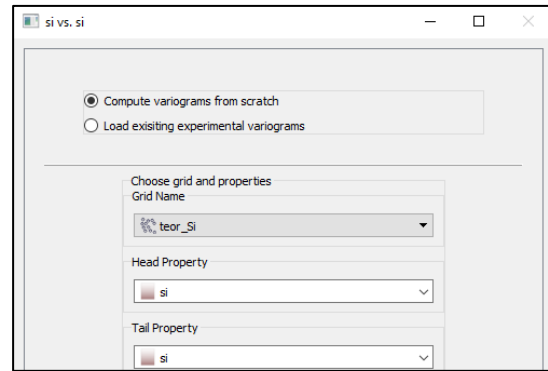


Figura 30: Carregamento dos dados para variograma.

Fonte: o Autor (2022).

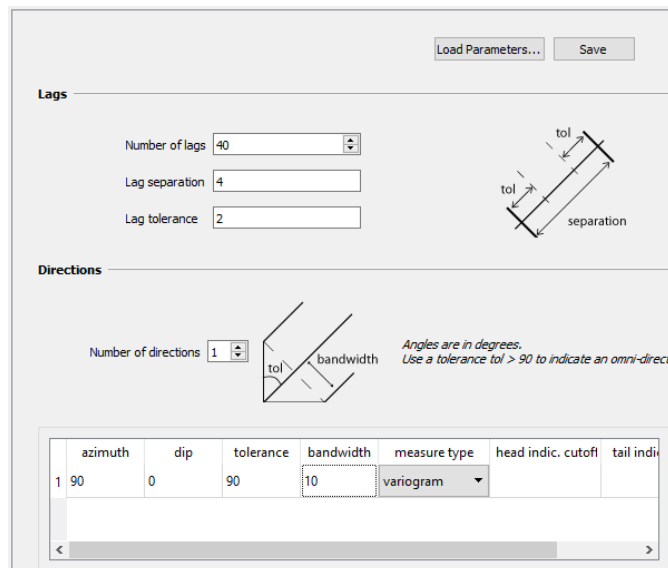


Figura 31: Parâmetros Variograma.

Fonte: o Autor (2022).

Após a geração do variograma, cada ponto apresentará um número de pares de amostras encontrados e é preciso modelar uma função que se adeque aos resultados da busca. Assim, distribuímos a variância inicial dos teores com uma contribuição arredondada de 0,66 para o *Sill* e 0,008 para o efeito pepita. Normalmente, quando a escala do gráfico permite, é traçada uma linha entre os três primeiros agrupamentos de pares do variograma e o efeito pepita é considerado onde ela interceptar o eixo das variâncias. Dando continuidade, seleciona-se o modelo que mais se aproxima ao agrupamento dos pares, seus *ranges* máximo, mínimo e ângulo, nesse caso, modelo de uma estrutura esférica, 40 minutos, 20 minutos e 90°

respectivamente (Figura 32). Não foi realizado modelamento na direção norte por não haver amostras, o alcance avaliado é apenas para a direção leste.

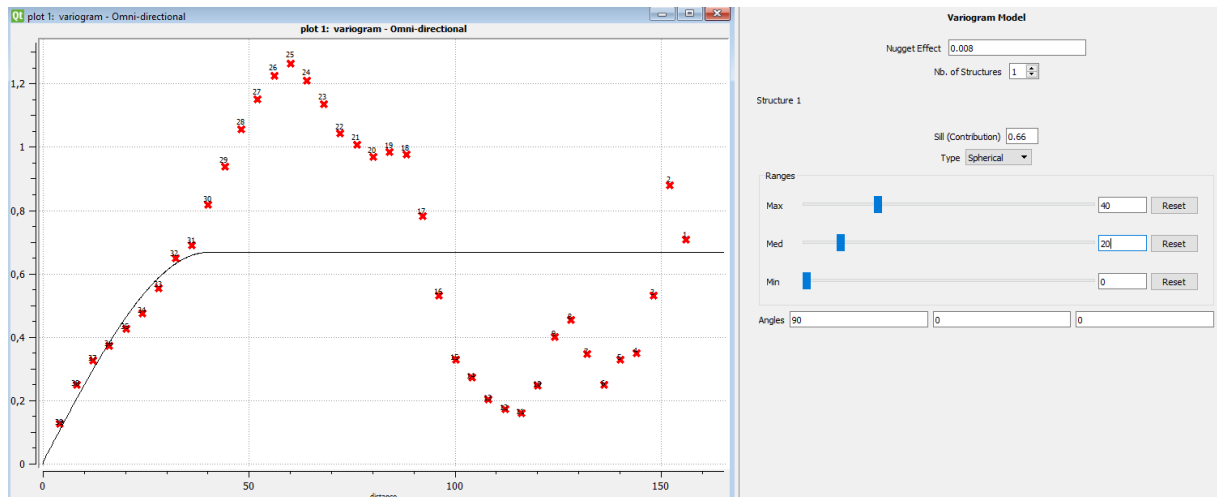


Figura 32: Modelamento da função variograma.

Fonte: o Autor (2021).

A próxima etapa é a utilização do algoritmo *Gammabar*. Ao abri-lo, o usuário é questionado sobre o arquivo de parâmetros que deseja utilizar, basta pressionar o botão *Enter* do teclado para criar um novo arquivo de parâmetros na pasta onde está o arquivo executável (Figura 33). O arquivo deve ser preenchido conforme os parâmetros do modelamento. Nesse caso, como existe apenas uma estrutura no variograma, as últimas duas linhas do arquivo serão excluídas.

Como o primeiro intervalo de tempo que se deseja descobrir a variância de extensão é 15 minutos, o primeiro dado referente ao tamanho do bloco em x é a metade disso, 7,5 minutos, y e z permanecem zero. A discretização é estabelecida como 1000 em x para que haja a convergência dos dados, y e z permanecem zero. Uma vez o arquivo preenchido, o *Prompt* de comando do sistema operacional deve ser aberto com o diretório correto carregado, é necessário o conhecimento do caminho do diretório em que o arquivo de execução *Gammabar* e o arquivo de parâmetros se encontram. Digitando o comando *gammabar.exe* e pressionando *Enter* duas vezes, o algoritmo calculará a variância para o intervalo de 7,5 minutos, 0,069652 neste caso (Figura 34).

Figura 33: Arquivo de parâmetros Gammabar.

Fonte: o Autor (2021).

Figura 34: Resultado da variância do EE para 7,5 min.

Fonte: o Autor (2021).

É necessária a repetição do procedimento de cálculo da variância para cada intervalo de tempo que se deseja descobrir a variância de extensão e para o intervalo referente a sua metade. Para isso, basta alterar o primeiro valor do arquivo de parâmetros e executar o algoritmo novamente. Uma vez obtidos todos os dados, é preciso aplicar a Equação 50, a variância de extensão do intervalo desejado é duas vezes a variância da metade do intervalo desejado em relação ao menor intervalo de coleta menos a variância do intervalo desejado em relação ao menor intervalo de coleta. Ainda, a fim de se obterem resultados dos erros com intervalo de confiança de 99,7%, retira-se a raiz da solução da equação e multiplica-se por 3 (três desvios padrões).

Na Tabela 7, estão os resultados dos erros de extensão em termos de variância e desvio absoluto em percentagem para um intervalo de confiança de 99,7%.

Tabela 7: Resultados do exemplo do cálculo computacional do erro da variância de extensão.

T(min)	σ^2 (·,T)	T/2=t (min)	σ^2 (·,t)	Var EXTENSÃO (·,T)	Erro de Extensão	Teor min	Teor max
15	0,1300	7,5	0,0697	0,00932	0,2896	4,020	4,600
30	0,2416	15	0,1300	0,01838	0,4067	3,903	4,717
45	0,3322	22,5	0,1878	0,04336	0,6247	3,685	4,935
60	0,3959	30	0,2416	0,08728	0,8863	3,424	5,196
120	0,5185	60	0,3959	0,27334	1,5685	2,742	5,878

180 0,5629 90 0,4726 0,38226 1,8548 2,455 6,165

Fonte: o Autor (2022).

Como é possível perceber analisando o Gráfico 12, quanto menor for o tempo de coleta entre os incrementos que formarão uma amostra, menor será a variância de extensão, ou seja, menor será o erro associado ao estender o valor de uma única amostra ao intervalo do lote que ela representa. Exemplo disso é o aumento da variância de extensão em mais de 300% quando o intervalo de coleta aumenta de 60 para 120 minutos.

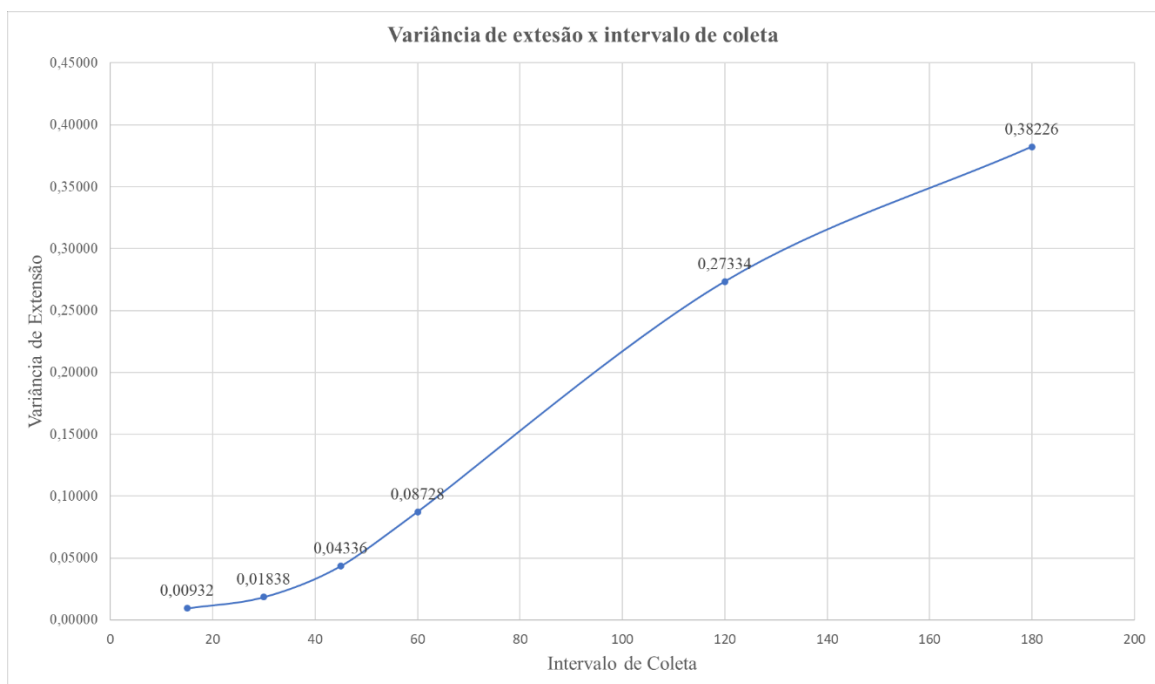


Gráfico 12: Exemplo variância de extensão x intervalo de coleta.

Fonte: o Autor (2021).

3.3. TESTE VARIOGRÁFICO

O teste variográfico tem por finalidade definir os erros de extensão cometidos durante a amostragem do transportador de correia na usina para diferentes intervalos de tempo (maiores do que o já praticado na coleta regularmente). O objetivo é apontar se a variância em torno da média das amostras coletadas no intervalo de tempo atual é condizente com os objetivos, tais como de reconciliação. No caso de um projeto de amostragem, esse teste pode ajudar a definir os intervalos de coleta, dimensionar amostradores e ainda orientar a equipe quanto aos aumentos dos erros em função de maiores intervalos de coleta. Partindo do princípio que todo o material é representado com relação à média e à variância do parâmetro estudado, antes

mesmo de nos preocuparmos com a variância, é preciso estimar o erro com relação à média. Ou seja, de acordo com a Teoria de Gy, só passamos ao cálculo da variância a partir do momento em que nos certificamos de que o viés foi eliminado.

Assim, de fato, após um teste variográficos, existem etapas de amostragem que visam à determinação do viés. As variâncias, isto é, as incertezas em torno da média que repercutem nas questões de reconciliação, assim como nas de controle de processo, serão tratadas em um segundo momento. Dessa forma, as diferenças com relação à média costumam ser maiores tratando-se de planejamento a curto prazo: semana a semana, mês a mês. À medida que a escala de tempo aumenta, planejamento a médio e longo prazo, as massas totais coletadas e analisadas aumentam (proporcionalmente ao número de incrementos), o que faz com que as diferenças de média fiquem menores. Em outros termos, as massas dos incrementos se aproximam, ainda que minimamente, da massa do lote original.

Normalmente, as médias dos teores dos materiais lavrados em um determinado período de tempo são comparadas, ou seja, a estimativa da média advinda do planejamento de lavra (amostras de perfuração, por exemplo) é comparada com as amostras coletadas na usina. Um dos primeiros fatores limitantes neste método de comparação se trata das diferenças de suporte amostral. Tanto quanto o método, as massas e a granulometria dos materiais coletados na mina são diferentes das pertencentes aos materiais coletados na usina. Em especial, a relação massa da amostra, quando comparada à massa do lote verdadeiro, é muito diferente na mina e na usina. O modo de preparo das amostras também difere muito. Por exemplo, na usina, diferente da frente de lavra, as amostras coletadas durante um turno ainda podem ser compostadas, o que significa que o número de incrementos tem grande impacto sobre a variância das amostras.

Como ilustração, é importante analisar os métodos de coleta de amostras na mina, que pode ser feito através de pó de perfuratriz, de testemunhos (cortados ao meio ou a um quarto) ou a partir de trincheiras ou canaletas. A chance de ferir a chamada Lei da Equiprobabilidade e, portanto, causar viés é grande no caso do pó de perfuratriz, porque pode ocorrer perda de material, segregação e agrupamento das partículas. Mesmo equipamentos que coletam o material durante a perfuração e façam o quarteamento automaticamente (em uma bandeja radial) são questionáveis do ponto de vista de viés e de variância do erro fundamental. Isto ocorre porque não há uma redução granulométrica e mistura prévias antes do quarteamento e pela perda de finos no processo de perfuração. Como visto anteriormente, isso pode ser tão problemático quanto realizar coletas de um material fragmentado caminhando sobre uma pilha utilizando uma pá. No caso de testemunhos, novamente, dada a irregularidade do cilindro,

heterogeneidade do minério e problemas de recuperação, podem ocorrer desvios maiores. Assim, a probabilidade de vieses e problemas de reprodutibilidade podem ser grandes, porém, são frequentemente ignorados.

Após essa preocupação primordial com o viés, a atenção é voltada para a representatividade da variância do material amostrado. De novo, aí surgirão questões relativas às diferenças de suporte, as quais estão relacionadas ao efeito de escala com respeito à heterogeneidade. Quando alterado o tamanho do bloco para as estimativas do teor, ou a distância entre os furos na bancada, ou os intervalos de tempo de coleta no transportador de correia, é alterando também o suporte amostral, e, conseqüentemente, mudado o valor da variância, sua representação na variância intrínseca do lote verdadeiro e nos erros. Se fosse possível analisar todo o material da bancada ou o lote que passa em um turno/dia na correia (lotes verdadeiros), seria possível acessar o valor de média e variância real da população que se quer representar. Entretanto, isso é operacionalmente impossível de ser realizado, assim, é natural que as variâncias medidas sempre sejam fortemente dependentes do suporte amostral. Parte dessa variância diz respeito a variância real do material e a outra parte pode ser um “inchaço” em função dos erros cometidos.

O teste variográfico tradicional em correia, geralmente, procura empregar o menor intervalo de tempo de coleta possível, em torno de 2 em 2 minutos a 4 em 4 minutos, durante um tempo total de coleta de duas a quatro horas. De forma resumida, este teste já foi descrito no Item 3.1, onde foi apresentada a dedução da fórmula do Erro de Extensão. Este teste é aconselhado, por exemplo, para fins de reconciliação através de prognosticação para modificações em metodologias e processos. O teste em si tem algumas implicações importantes sobre as estimativas das incertezas que a maioria das pessoas desconhece, como sua aplicação para obtenção do erro de extensão. O seu planejamento e os seus resultados dependem da forma como as amostras são coletadas na mina e na usina, bem como se há uma pilha de blendagem. Simulações geoestatísticas podem ser utilizadas para reproduzir a variabilidade de teores em pilhas e processos de amostragem de fluxo contínuo, sendo possível emular testes variográficos (Marques, 2013).

Nota-se que há possibilidade de aplicar testes variográficos no ROM, no concentrado ou no rejeito. Isso dependerá do *top size* do material amostrado, da existência ou não de amostradores automáticos, da logística ou suporte por parte da empresa no que tange à coleta, transporte, preparação e análise das amostras. A coleta visando à medida do teor do ROM deve ser efetuada nos locais onde a granulometria for menor, no último estágio de britagem ou após

a moagem por exemplo. Entretanto, a amostragem pós moagem pode ser mais complexa em função do material estar em forma de polpa, oferecendo especial dificuldade para garantia da representatividade na etapa de tomada das amostras.

Ainda, quanto à problemática da granulometria, coletas em determinados pontos do circuito, como na alimentação de um moinho semiautógeno ou de um britador (onde os tamanhos são maiores), não contam com metodologias de razoável confiabilidade ou equipamentos adequados para fazê-las. Existem alguns guias empíricos pouco detalhados, como o descrito no livro de Napier-Munn & Morrell & Morrison (1996), porém, quando a representatividade com respeito ao lote verdadeiro de algumas das metodologias de tomadas de amostras descritas nesses trabalhos é analisada, é verificado facilmente que não é atingido o nível de representatividade adequado. Historicamente, muito trabalho e tempo foram dedicados a esse tipo de amostragem (visando à granulometria de cabeça), pois ela auxilia a verificar o dimensionamento e as variáveis operacionais dos equipamentos de cominuição, assim como os efeitos disso sobre o processo de liberação e, por consequência, de recuperação metalúrgica.

Dessa maneira, embora seja necessário proceder com testes variográficos para dimensionar ou adequar amostradores automáticos e evidenciar os efeitos negativos da amostragem mal conduzida, na prática, existem muitas dificuldades ligadas: (1) ao desafio tecnológico para coleta de material ROM, (2) à falta de guias completos na literatura e (3) à aceitação da sua importância por parte da indústria.

3.3.1. Teste Variográfico Personalizado

A elaboração de um teste variográfico depende de diversas questões, como o objetivo do teste, tipo de minério, heterogeneidade do material, etapa onde ocorrerá a coleta, entre outros. Cada situação demanda um estudo aprofundado e um teste personalizado para que os melhores resultados sejam obtidos. Um teste que objetiva verificar as imprecisões cometidas em “amostragem” de pilhas (a qual é proibida pela teoria de Gy, mas ainda de comum prática) pode ser elaborado conforme será descrito na sequência.

Em testes desse tipo, as maiores incertezas residem na amostragem dos elementos ou componentes menores, por isso, escolhem-se dois parâmetros como referência: um elemento ou fase mineral abundante e um elemento menor ou contaminante. Geralmente, a empresa

mineradora, em algum momento, tende a fazer amostragem em pilhas por razão de armazenagem do minério, presença de pilhas pulmão, blendagem ou transporte por vagões de trem, caminhões ou navios. Por conseguinte, infelizmente, a empresa cliente ou compradora costuma fazer o mesmo tipo de prática não indicada. Para a transação, ambas entram em um acordo (contrato) sobre o valor mínimo ou médio do elemento de interesse e sobre o valor máximo dos contaminantes. Desconsiderando as melhores práticas, ignoram o que seja um teste variográfico ou erro de extensão. Assim, é comum encontrar divergências vindo de ambas as partes que levam a multas contratuais e até mesmo a processos judiciais.

Aqui, para que a compreensão seja facilitada, é proposto um exemplo semelhante ao estudo de caso que será apresentado ainda neste trabalho, no Capítulo 5. Nesse exemplo, no cliente, o procedimento de amostragem é realizado em pilhas, o material recebido da mineradora é basculado pelos caminhões (de 15t), uma a duas pilhas por vez, em um pátio específico. Após 20 caminhões, estas sofrem uma mistura e divisão por retroescavadeira em pilhas de 30t, em cada pilha são retirados 30 incrementos de 0,5 kg com pá e, após completar 300 kg, acontece uma divisão no “estilo cone e cruzeta” resultando em pilhas cada vez menores: de 300 kg até cerca de 5 kg. Ao final, uma das quatro fatias é selecionada com uma pá, consistindo na amostra. A cada vinte amostras, retira-se a outra fatia como duplicata. Esse processo provoca claramente uma imprecisão.

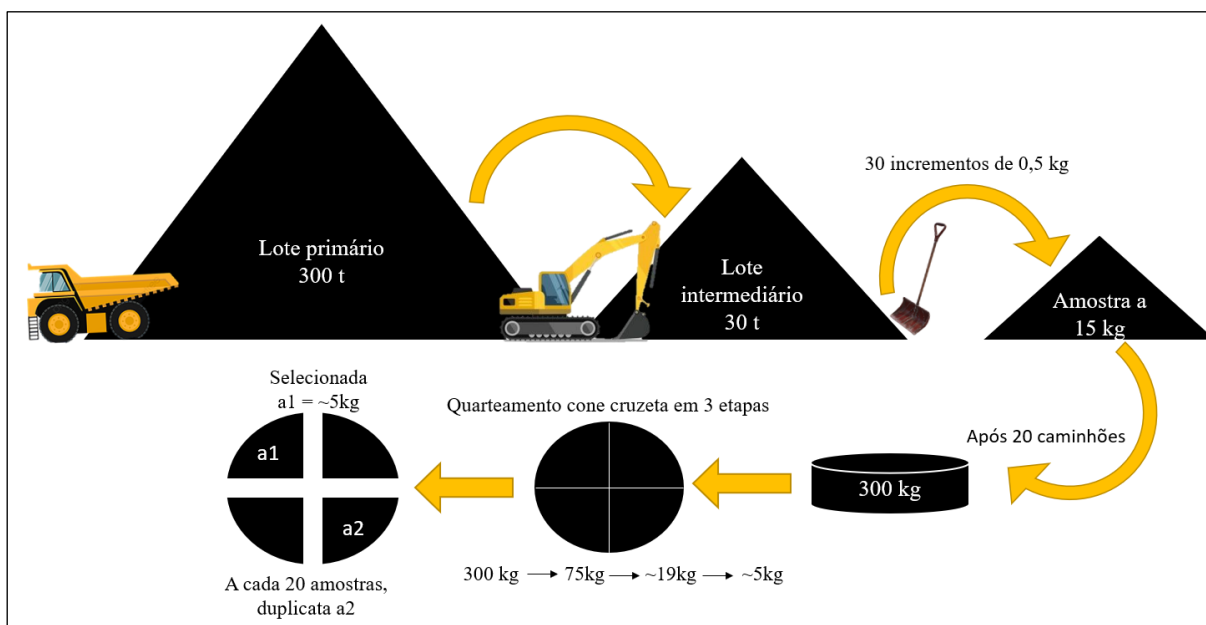


Figura 35: Esquema representando a amostragem praticada em pilhar no cliente.

Fonte: o Autor (2022).

Assim, um teste variográfico customizado pode ser realizado escolhendo-se primeiramente três pilhas, denominadas de Lotes, de aproximadamente 100 toneladas cada. As pilhas alimentarão, uma de cada vez, um silo. O material do silo será descarregado numa correia transportadora a uma taxa aproximada de 30 t/h. Neste caso, não havendo amostrador automatizado, a correia transportadora deve ser parada e todo o material coletado por um gabarito.

O gabarito é formado por duas placas metálicas para delimitar cerca de 2 metros de uma pequena correia com 40 cm de largura. Sugere-se que o material coletado entre as placas corresponda a cerca de 10 kg por tomada (base úmida). O tempo de parada pode ser da ordem de 30 em 30 minutos para coletar amostras com o gabarito. Dentro deste procedimento, devem ser coletadas amostras em duplicatas (aconselha-se 25% de taxa de inserção, cinco vezes maior do que o teste convencional), além de brancos (material semelhante ao minério e sem o elemento menor analisado) e padrões (material de teores conhecidos) inseridos durante a preparação das amostras (Abzalov, 2016). Estas amostras devem ser cuidadosamente embaladas em separado e levadas ao laboratório para homogeneização. Após esta etapa, as amostras devem ser peneiradas em malha igual à APF (abertura de posição fechada) do britador. O retido deve ser britado no menor tamanho possível. Após realizar a homogeneização e/ou quarteamento no equipamento Jones ou divisor rotativo, a alíquota final deve possuir massa adequada para a análise em laboratório de acordo com o tipo de minério.

Calcula-se o número de amostras que serão coletadas para representar o lote primário de 300t do cliente, ou seja, aproximadamente 20 amostras de 10 kg cada no total. Ainda, se recomenda a tomada de 10 amostras chamadas de “sequenciais” (simulam a condição de tempo de coleta “zero” minutos) a cada 2 tomadas, totalizando relação de 300kg de 300t.

Este teste leva cerca de sete dias para ser realizado no que tange à coleta das amostras. Geralmente, são necessários cerca de 30 dias para preparar e analisar todas as amostras e mais 30 dias para confirmar os resultados, repetindo o processo de preparação e análise das amostras desde o princípio. O esquema simplificado do teste pode ser observado na Figura 36.

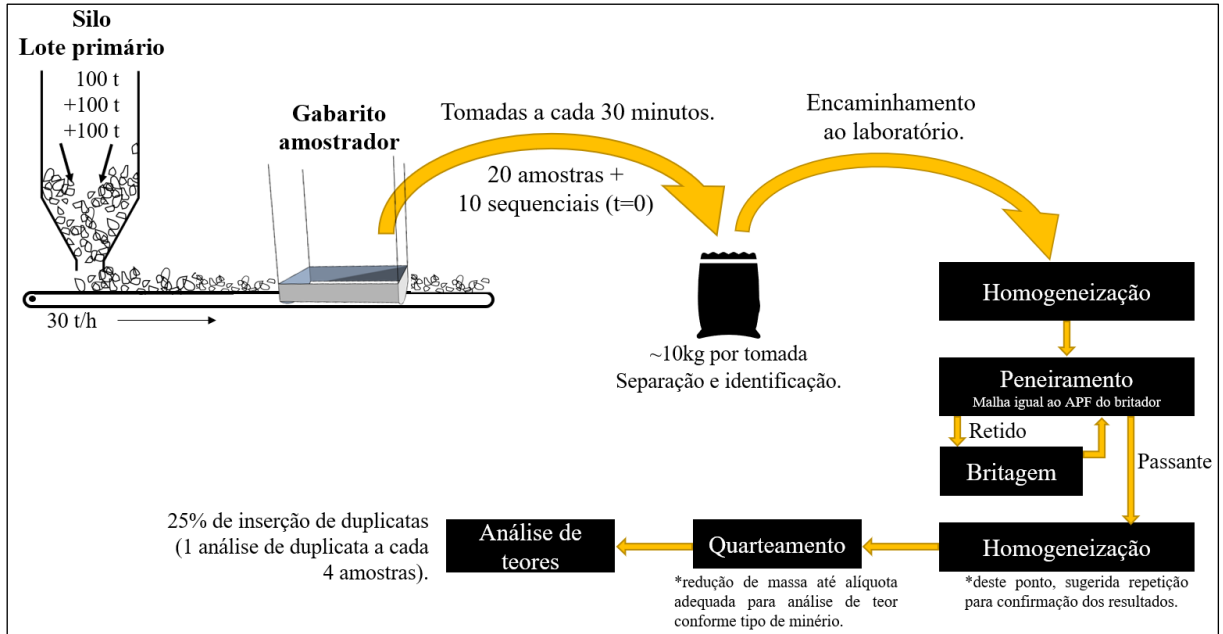


Figura 36: Fluxograma do teste variográficos personalizado.

Fonte: o Autor (2022).

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

Uma vez apresentada a variância de extensão e a maneira com que os dados podem ser adquiridos de forma personalizada para estimá-la (teste variográfico), é necessário debater como esse erro deve ser agregado ao modelo discreto proposto por Gy. Dessa forma, neste capítulo, serão debatidas as questões a respeito da importância e das problemáticas que envolvem a tentativa de união entre os modelos contínuo e discreto na amostragem mineral, para isso, serão utilizados exemplos teóricos com variadas hipóteses.

4.1.RELAÇÃO ENTRE MODELO DISCRETO E CONTÍNUO

Considerando todos os tipos de erros de amostragem de Gy, segregação e agrupamento, de seleção, erros de materialização e o erro fundamental, segundo a sua teoria, somente o erro fundamental poderia ser calculado diretamente, e sendo possível apenas quando minimizado o viés. No entanto, se não há forma de acessar estes erros por cálculos ou testes, também não existe certeza se o valor do erro fundamental, ou mesmo dos erros de precisão da amostragem, estão corretos. Conforme é visto na Figura 37, estes erros vão se somando e resultam no erro global da estimativa. A problemática é que, durante a preparação, não somente na tomada primária, ocorrem erros de seleção, de segregação e de agrupamento, pois não existem misturas perfeitas nem divisões perfeitas, principalmente no caso de minérios. Por isso, muitas vezes, a fórmula do erro fundamental de Gy é questionada, tal como no artigo Geelhoed (2011) que afirma que ela não é acurada.

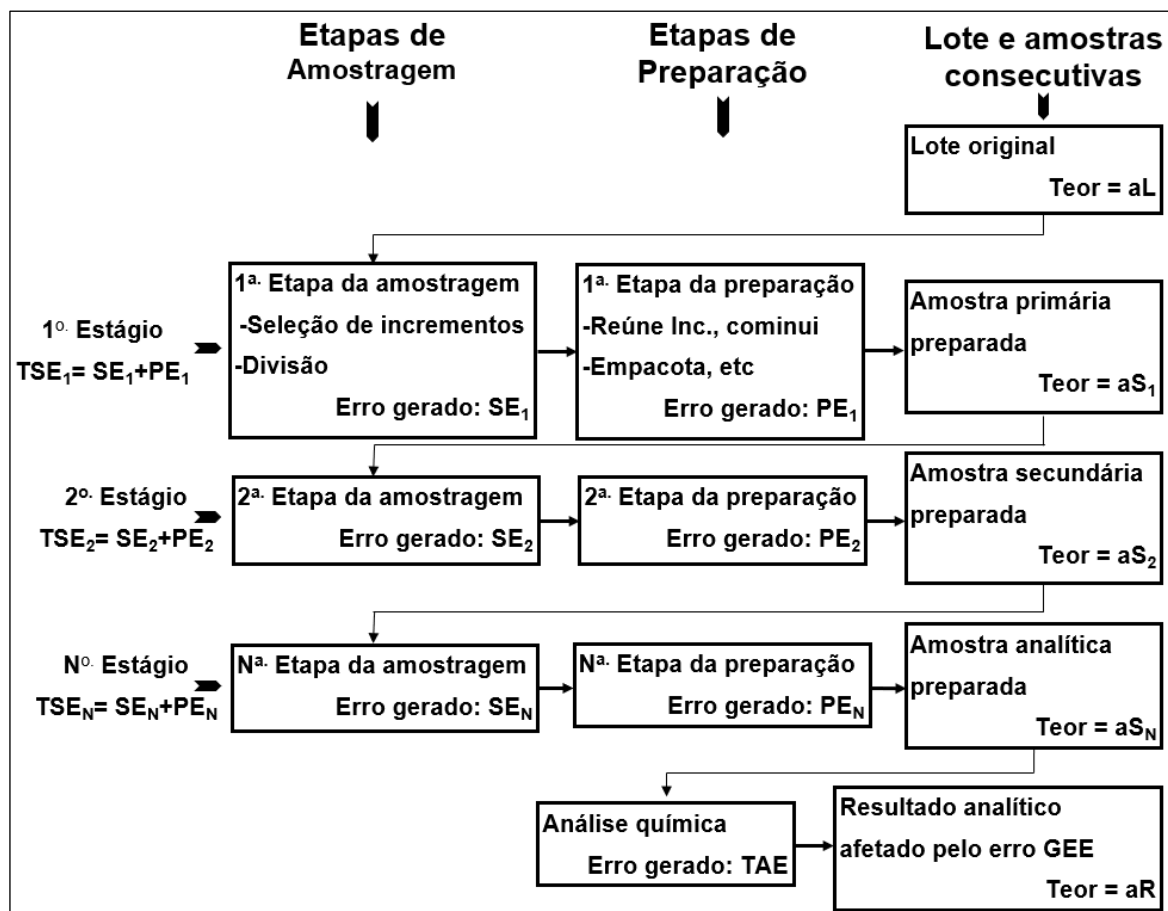


Figura 37: Somatório de erros em relação a estimativa de teor na preparação de uma amostra.

Fonte: modificada, Pitard (1993).

De fato, é válido observar com cuidado a fórmula de Gy para a variância do erro fundamental novamente, agora, tanto seu formato original (Equação 31) quanto o calibrado. Quando, no protocolo, não há redução de massa, calcula-se a variância apenas para fins gráficos, conforme a Equação 51 (Minnitt et. al., 2007). A partir desta, pode ser deduzida a Equação 52. Depois, são realizadas simplificações pertinentes (Equação 53). Na Equação 54, tomamos α como 3-b. A Equação 55 retorna com o termo K, substituído na fórmula. Assim, a Equação 56 é o rearranjo dos termos para calibração, a aplicação do algoritmo natural resulta, por fim, na equação de Gy capaz de ser calibrada com testes de heterogeneidade (Equação 57). Porém, a calibração só é possível através da obtenção dos parâmetros K e α do minério, obtidos por meio do resultado de testes heterogeneidade com massas quase sempre pouco representativas frente ao lote.

$$s_{(FE)}^2 = \left(\frac{1}{M_S}\right) \cdot K \cdot d_N^\alpha \quad \text{Equação 51}$$

$$s_{(FE)}^2 = \frac{c \cdot f \cdot g \cdot \left(\frac{dl}{d_N}\right)^b \cdot d_N^3}{M_S} \quad \text{Equação 52}$$

$$s_{(FE)}^2 = \frac{c \cdot f \cdot g \cdot dl^b \cdot d_N^{(3-b)}}{M_S} \quad \text{Equação 53}$$

$$s_{(FE)}^2 = \frac{c \cdot f \cdot g \cdot dl^{(3-\alpha)} \cdot d_N^\alpha}{M_S} \quad \text{Equação 54}$$

$$s_{(FE)}^2 = \frac{K d_N^\alpha}{M_S} \quad \text{Equação 55}$$

$$s_{(FE)}^2 \cdot M_S = K d_N^\alpha \quad \text{Equação 56}$$

$$\text{Ln}(s_{(FE)}^2 \cdot M_S) = \alpha \cdot \text{Ln}(d_N) + \text{Ln}(K) \quad \text{Equação 57}$$

Onde:

K: c.f.g.d_l^{3-α}

Uma das limitações mais importantes da fórmula de Gy, calibrada ou não, quando considerada a redução de massa da amostra em relação ao lote, é que, quando $M_L \geq 10 M_S$, o termo “ $1/M_L$ ” é tão pequeno que se torna insignificante. Quando taxas de reduções de massa grandes são aplicadas, a variância do erro fundamental se torna grande. Será demonstrado, no Capítulo 4, através de alguns exemplos quali e quantitativos, as consequências dessa limitação. De qualquer forma, esclarecendo, é muito difícil, quase impraticável, calcular a variância do erro fundamental pela fórmula de Gy de modo a obter a representatividade de um saco de amostras com alguns quilos diante de toda a massa de uma bancada. Esta fórmula é adequada para desenhos de protocolo amostral apenas de preparação da amostra em laboratório, a qual denominamos de lote zero dimensional. Não é possível estabelecer uma relação entre este lote e o lote verdadeiro tridimensional que se quer representar (uma bancada, um depósito, o ROM que alimenta a usina), ou seja, entre as suas massas. De fato, não existe como calcular a correlação ou determinar continuidades por meio da fórmula do erro fundamental de Gy no

modelo discreto mostrado na Figura 38. Por outro lado, por meio da Geoestatística (variografia) é possível estimar os erros relacionados à amostragem e o “inchaço” do efeito pepita, uma vez que as amostras estariam correlacionadas no espaço e no tempo.

Ainda, na fórmula de Gy não calibrada, além do valor do erro poder ser bastante enviesado por tratar-se de constantes tabeladas e/ou calculadas, existe a limitação da obrigação do uso do teor do mineral na equação original, e não do elemento químico. No caso do carvão, por exemplo, mede-se o teor de cinzas, as quais representam o conjunto de óxidos gerados na combustão a partir dos minerais. Estes minerais podem ser argilas, quartzo, carbonato, dentre outros. É praticamente impossível realizar uma estequiometria para uso na fórmula de Gy. Problemas dessa natureza também ocorrem com os elementos contaminantes na maioria dos minérios, por exemplo: o magnésio, no minério de ferro, pode aparecer em mais de 3 tipos de minerais e, inclusive, substituindo outro íon em qualquer outro mineral mais abundante, dando a uma estequiometria elevado grau de complexidade. Já os testes de heterogeneidade permitem calibrar as constantes com base no teor do elemento químico. No Capítulo 4, assim como no Capítulo 5, serão apresentados exemplos práticos dessas limitações e seus impactos. O Capítulo 5 constitui-se de um estudo de caso, ou seja, um trabalho de aplicação da teoria de Gy e, em especial, do modelo contínuo para demonstração da sua importância.

Testes de heterogeneidade, ou mesmo testes variográficos, deveriam ser realizados rotineiramente, porém o mais comum é que não sejam executados pela empresa ou nem mesmo conhecidos. Uma desvantagem ou limitação desses testes consta no fato de que ainda dizem respeito a uma massa muito pequena diante do depósito todo e, portanto, pode ser questionada a sua validade em relação ao ciclo de vida da mina. Por exemplo, um teste de heterogeneidade geralmente usa cerca de 500 kg (Chierigati, 2008), (François-Bongarçon, 2002). Em um teste variográfico, mesmo que executado por quatro horas, as massas ainda serão pequenas com relação a todo o depósito.

Observando com cuidado a Figura 38, com respeito ao modelo discreto, nota-se que o erro devido a pequenas variações na heterogeneidade de curto alcance significa mudanças na heterogeneidade de constituição do minério em intervalos muito pequenos (espaciais ou temporais). Tal como visto anteriormente no protocolo de preparação de testemunho, dividir o testemunho na metade ou em $\frac{1}{4}$ é algo disparatado, visto que nenhum testemunho tem geometria regular ou perfeita. Ao fazer isso, para a maior parte das amostras de um depósito, se está incorrendo também em um erro de seleção contínua para heterogeneidade de curto alcance, como pode ser visto no modelo contínuo.

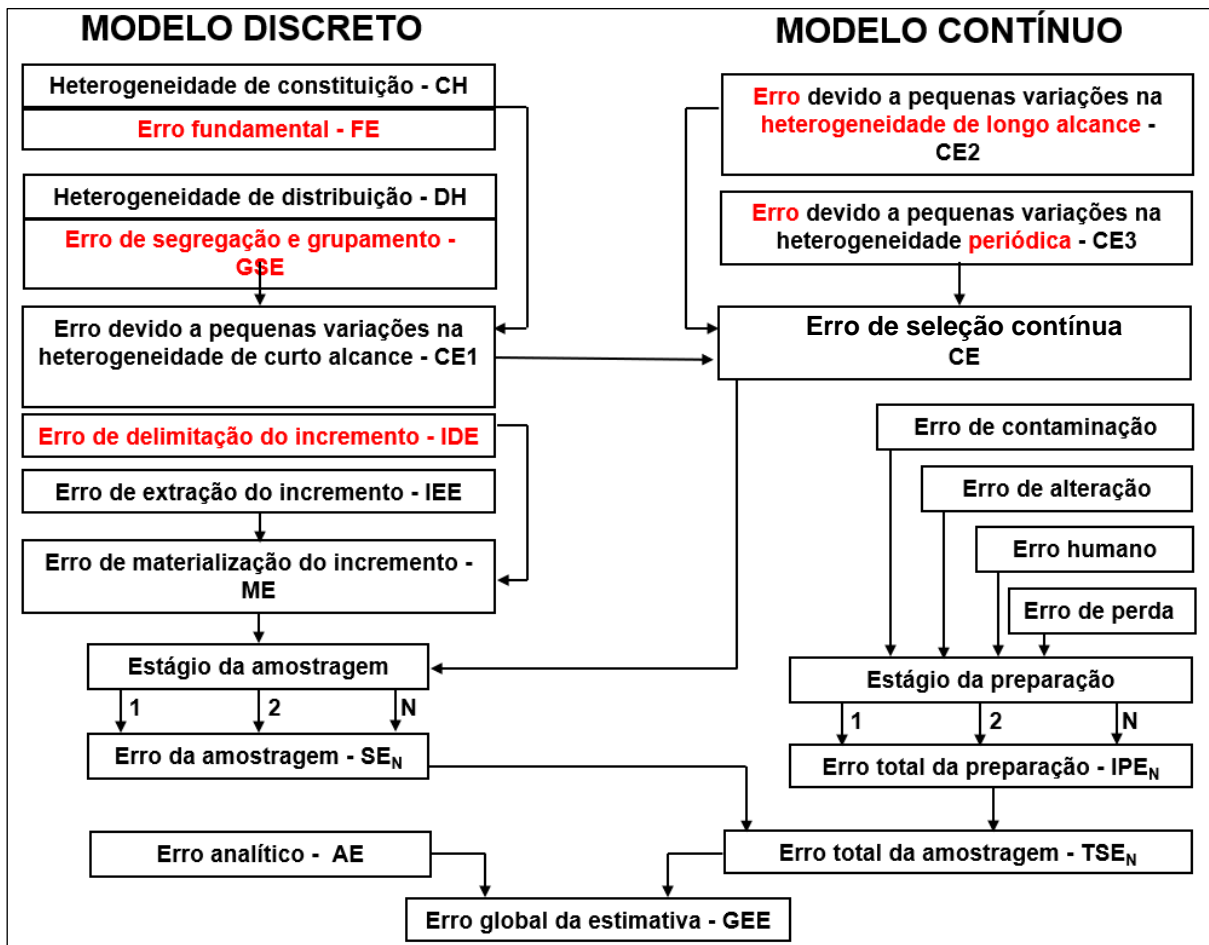


Figura 38: Esquema entre os erros de amostragem dos modelos discreto e contínuo.

Fonte: modificada, Pitard (1993).

O fato de também haver ou não uma blendagem (tal qual sua qualidade) interfere nos erros de longo alcance periódicos (CE3). Por exemplo, quando não há blendagem, o material pode segregar de tal modo que, em um turno, predominam teores baixos e elevados no outro, afetando sobremaneira as operações de concentração. Frequências de coleta sobre a correia transportadora devem acompanhar as mudanças de massa ao longo da correia, as mudanças de massa dos lotes verdadeiros, bem como a mudança nas heterogeneidades de constituição (CE2). Ambos tipos de erros CE3 e CE2 levarão a um erro de seleção contínua (CE), incluindo erros de viés, de precisão e de determinação da heterogeneidade intrínseca do lote. Recordar-se que o minério é constituído por fragmentos de diferentes formas, tamanhos, densidades e teores. Quando comparados fragmento contra fragmento, é visualizada a heterogeneidade do minério. Por exemplo, o responsável pela amostragem, seguindo o protocolo, realiza o quarteamento da

amostra através de um equipamento “cone e cruzeta”. Disto, resultam 4 “fatias”, das quais ele escolherá duas diagonalmente opostas, como um par de duplicatas a ser analisado. Ao verificar que são muito diferentes os teores, ele poderá ter a impressão de que a mistura anterior ao quarteramento não foi adequada, ou, o tamanho do material e a massa da amostra também não estão adequados. Assim, ele decide pegar sempre as quatro porções e descobre que há uma grande diferença de teores entre elas. Nessa situação, é possível concluir que esse minério é bastante heterogêneo e que, ao aumentar o número de incrementos, a variância não diminuiu, mas aumentou. Novamente, parte dessa variância é da heterogeneidade intrínseca do minério, e parte dela dos erros de amostragem cometidos, os quais possuem a tendência de aumentar na proporção em que o material é mais heterogêneo.

Em realidade, discute-se aqui uma questão que perpassa a mudança de suporte. Na usina, a título de exemplo, também é comum executar um teste chamado de alta frequência. Este consiste em aumentar o número de incrementos coletados e comparar as médias deste teste com os procedimentos de rotina (onde o número de incrementos é bem menor). Sendo as médias muito diferentes nos resultados, há a inclinação natural para a conclusão de que a média que mais se aproxima da média do lote verdadeiro é aquela que possui maior número de incrementos, fato que não é necessariamente verdadeiro. Isso ocorre pois o somatório total das massas das amostras ainda assim será menor que a massa do lote verdadeiro. O cálculo dessa média e dessa variância não tem relação com o erro de extensão pois não fazem uso da variografia.

Para que as explanações anteriores fiquem expostas de maneira clara, na Figura 39 foram esquematizadas (graficadas em azul) as distribuições dos resultados de 4 distintas rotinas de amostragem do teor de um lote L de média e distribuição verdadeiras conhecidas (graficadas em verde). Cada situação oferece uma distribuição de dados diferente como resultado. A primeira diz respeito a uma amostragem ideal, enquanto as outras três amostragens são problemáticas, por motivos de erros de amostragem de Gy ou quantidade mássica de incrementos insuficiente, tornando suas variâncias diferentes da do lote original e contribuindo na geração de erros. É sabido que, na prática, a média e a variância do lote quase nunca são conhecidas, a esquematização apenas sugere que resultados de amostragens enviesadas e/ou mal conduzidas podem induzir a erros de interpretação. Os gráficos de pizza completos representam a IH_L (Heterogeneidade Intrínseca do Lote) verdadeira (verde) e as calculadas (azuis).

Seguem as 4 descrições comentadas:

- a. **Variância e média da amostragem aproximadamente iguais às do lote (amostragem ideal):** nesse caso, o somatório das massas dos incrementos se aproxima da massa do lote (massa do lote pequena), as taxas de reduções estão entre 1 para 2 e 1 para 5 e existe o seguimento de todas as regras de Gy: mistura, quarteamento e *top size* adequados (controle facilitado pelas massas pequenas). Ainda, a tomada da amostra é feita sem ferir a Lei da Equiprobabilidade, preferencialmente, no caso de ROM, após o último estágio da britagem. RESULTADO: a média das amostras é aproximadamente a média do lote ($\bar{x}_L \cong \bar{x}_a$), ou seja, sem vieses. Ainda, a alíquota da variância que diz respeito ao erro de amostragem é minimizada, ou seja, quando subtraída a variância do erro (duplicatas) da variância das amostras, o resultado é aproximadamente o IH_L verdadeiro ($\sigma^2_L \cong \sigma^2_a$).
- b. **Variância da amostragem maior que a do lote e médias diferentes:** amostragem incorreta, há um alargamento na distribuição da amostragem ($\sigma^2_L < \sigma^2_b$), erros de amostragem incham a variância e “camuflam” o verdadeiro valor de IH_L (parecendo menor). A média das amostras é maior que a média do lote ($\bar{x}_L \neq \bar{x}_b$), ocorre viés.
- c. **Variância da amostragem muito maior que a do lote e médias aproximadas:** amostragem incorreta, há um grande alargamento na distribuição da amostragem ($\sigma^2_L \ll \sigma^2_c$), erros de amostragem incham a variância e “camuflam” o verdadeiro valor de IH_L , tornando-o, de forma aparente, significativamente menor que o verdadeiro. Os resultados com médias aproximadas entre amostras e lote podem levar a crer no sucesso do protocolo amostral ($\bar{x}_L \cong \bar{x}_c$).
- d. **Variância da amostragem menor que a do lote e médias muito diferentes:** amostragem incorreta, há uma subestimativa da variância real do lote ($\sigma^2_L > \sigma^2_d$), levando a crer que, subtraindo a variância do erro da variância das amostras novamente, a heterogeneidade intrínseca do material é muito pequena. Nesse caso, o viés superestima a média real do lote.

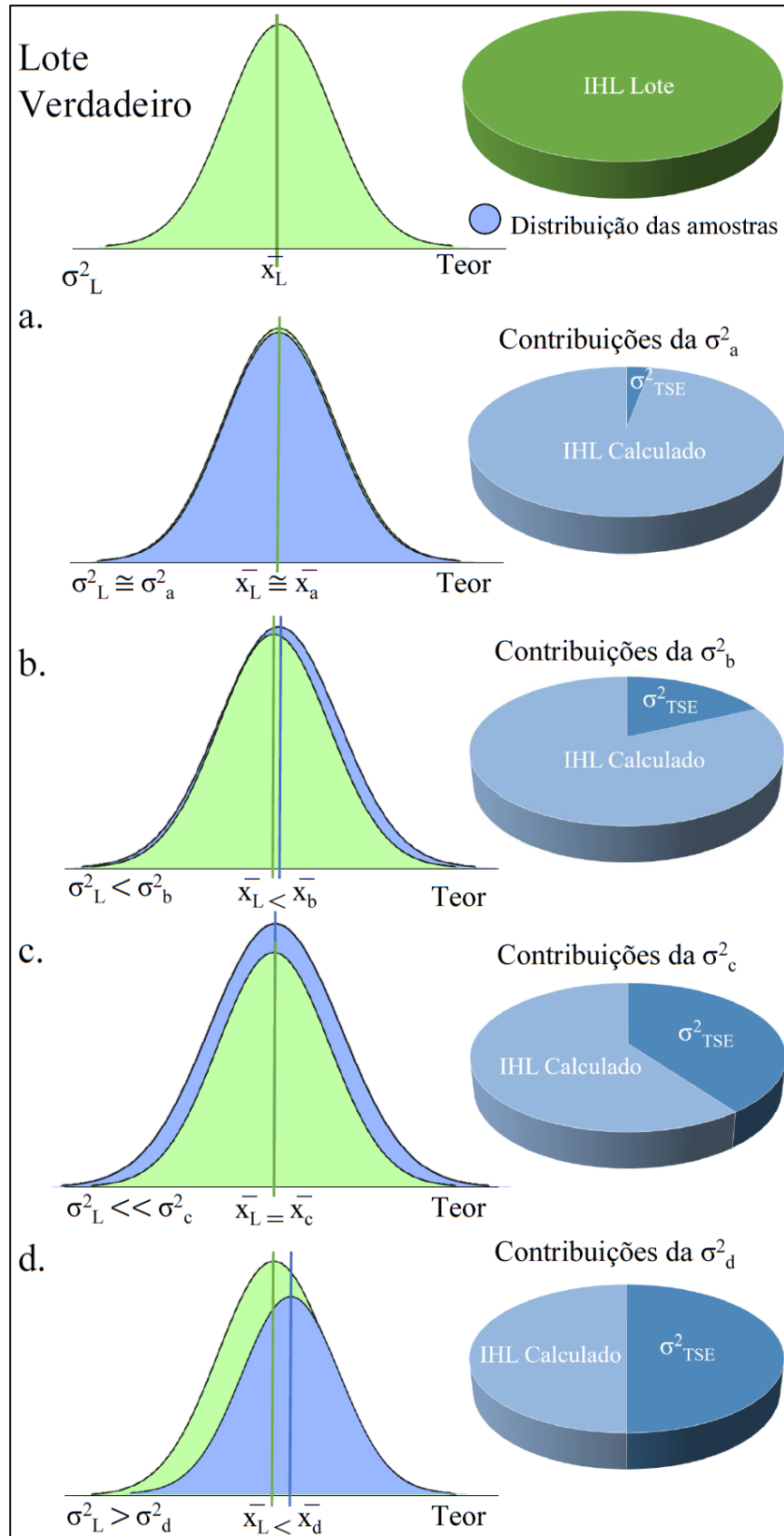


Figura 39: Distribuições de resultados de diferentes amostragens frente a um lote.

Fonte: o Autor (2022).

Por fim, é importante ressaltar outro aspecto notado com estreita ligação com os erros da teoria da amostragem. Os resultados das análises químicas de um material proveniente diretamente do ROM, que passa por uma britagem de boa qualidade e posteriormente por uma blendagem também de qualidade, não sofrem significativas alterações com a variação do número de tomadas de incrementos, os resultados das análises de cada incremento são muito semelhantes nessa situação. O fenômeno é observado por conta da mistura e divisão eficazes que o minério sofreu, elas diminuem valor de DH_L ao seu mínimo, próximo da nulidade, e tornam a relação massa da amostra diante da massa do lote um parâmetro mais determinante no QAQC que a própria diferença entre os pares de duplicatas.

A fim de ilustrar esse fenômeno e a importância da mistura e quarteamento, em especial na amostragem de ROM, retoma-se o exemplo anterior. Considerando uma amostragem de um material heterogêneo, mas com britagem e homogeneização eficientes, em correia pós britagem, com mistura das amostras em lona e, na última etapa do protocolo amostral, com quarteamento por “cone cruzeta”, chamado de “A” na Figura 40. Esse método de mistura e quarteamento não é preciso e oferece outras desvantagens além da falta de automaticidade. Pois, quando comparado com o divisor de rifles (Jones) e com divisor rotatório, logo se verifica que o número de incrementos deste último será maior, e assim, tendo sido feita uma boa mistura, a precisão também. Pois, no Jones simples, existe apenas a possibilidade de obtenção de 2 incrementos (2 bandejas) ao fim da preparação e, em um divisor rotativo de 6 caçambas, considerando a mesma massa de amostra, ao final, ao invés de 2 incrementos, é possível obter 6. Já o “quarteamento de pilha por cruzeta” oferece 2 fatias diagonalmente opostas, ainda assim, no exemplo, se analisadas as quatro fatias (amostra esgotada em laboratório), e constatada a variância maior do que com apenas 2, esta variância não é inchada devido aos erros qualitativos, mas por que a heterogeneidade intrínseca da amostra foi acessada, pois a blendagem era efetiva. Analogamente, o mesmo ocorre caso analisadas as 6 amostras do divisor rotativo.

Além disso, outra situação hipotética relevante na compreensão do fenômeno será descrita, sendo o caso extremo oposto do anterior, chamado “B”, considerando uma pilha segregada como um lote que se quer amostrar, podendo ser de estoque, pulmão, alimentação de ROM “online” (sem blendagem) ou como material estocado em silos. De início, a decisão para o local da tomada da amostra já é um desafio, embora a principal dificuldade está na falta de homogeneização. No caso da presença de um silo para estocagem do material passante de hidrociclones em regime turbulento, assim como em pilhas, também ocorrerá elevada segregação no interior da estrutura. A variância dos pares de duplicatas do exemplo “A” será

menor que a variância do exemplo “B”. Entretanto, toda a descrição anterior do número final de amostras é válida para essa situação também, ou seja, a variância dos pares de duplicatas será significativa para determinar a heterogeneidade e a influência provocada pelos erros na elevação da própria variância dos pares. Assim, uma heterogeneidade grande “beneficia” o aumento da variância dos erros de amostragem. Coletar mais amostras com relação ao lote não garante que o erro diminuirá, especialmente quando verificamos que, de novo, a massa do lote é muito maior que o somatório das massas das amostras.

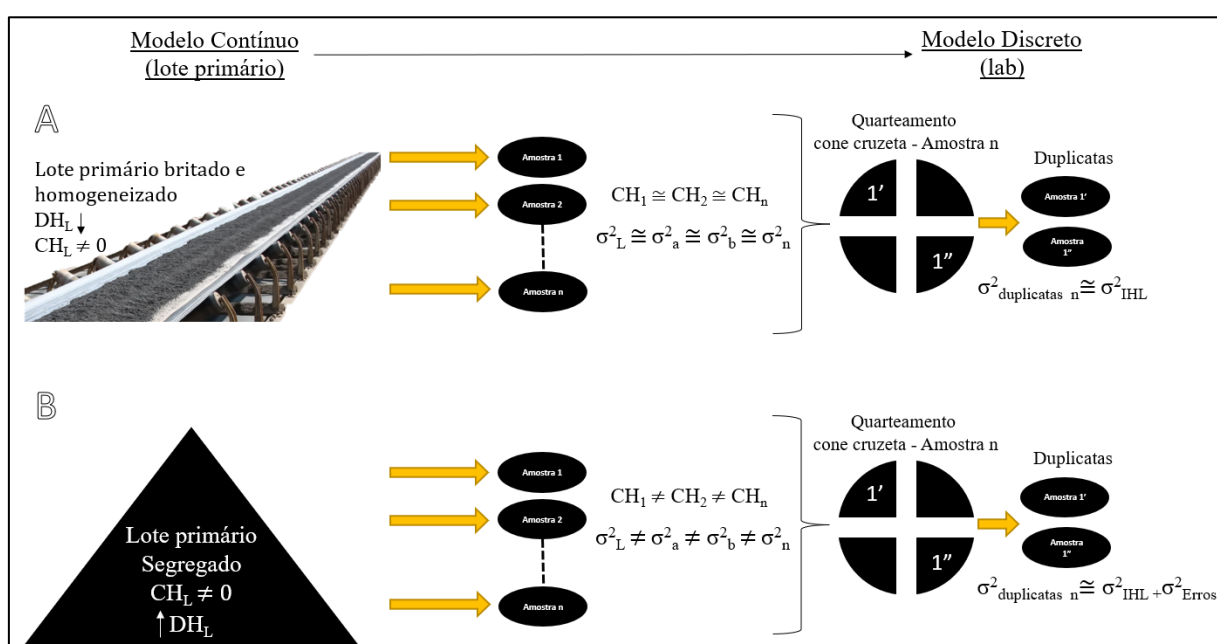


Figura 40: Comparativo das variâncias dos pares de duplicatas finais em diferentes protocolos amostrais.

Fonte: o Autor (2022).

Concluindo, uma boa blendagem tem impacto direto nos resultados de QAQC, nesse caso, a variância das duplicatas finais será condizente com a heterogeneidade intrínseca do lote e a homogeneização do ROM será capaz até mesmo de atenuar erros de amostragem. Assim, nesses moldes, conforme a Fórmula de Gy (Equação 31), a relação massa da amostra/massa do lote (que possui estreita relação com o erro de extensão) ganha maior importância que a variância das duplicatas. Já um material demasiadamente heterogêneo e segregado agrava erros de amostragem e, ao final, suas duplicatas possuirão uma variância composta não só pela heterogeneidade intrínseca do lote, mas também pela a variância dos erros qualitativos. Ainda,

devido a variações no DH_L , CH_L e nas massas do material ao longo do tempo, o aumento do número de incrementos não será garantia de que a variância das duplicatas se aproximará da variância intrínseca do lote.

4.2 ANÁLISE CRÍTICA DO TESTE VARIOGRÁFICO E IMPLICAÇÕES NO ERRO DE EXTENSÃO

Para que seja possível debater de forma consistente e didática a utilização do teste variográfico na definição do erro de extensão, um exemplo numérico simplista será estudado. A partir das metodologias estudadas e da discussão das problemáticas propostas no exemplo, uma análise será realizada. Ao fim, será esboçada a proposta de um novo protocolo QAQC e retomadas as considerações que dizem respeito à insuficiência do modelo proposto por Gy.

Inicialmente, para demonstração do que foi descrito anteriormente, é definido um exemplo constituído inicialmente por uma matriz de 16 células (4x4) representando teores médios do contaminante Óxido de Magnésio (MgO) nos lotes de minério de ferro que alimentarão uma usina de 30 em 30 minutos, totalizando um dia de trabalho de 8h. O lote possui teores entre 0 e 1% do contaminante para uma taxa de 2.000 t/h (Tabela 8). A massa de cada amostra quarteada que chega no laboratório de análises é de 0,5 kg. Para a melhor eficiência no beneficiamento, a usina exige que o teor de MgO no minério nunca ultrapasse 0,5% (Moraes, 2020).

Tabela 8: Matriz de teores colorida 4x4 (lotes 30 min).

0,5	1	0,5	0,5
0	0	1	0,5
0,5	0	0,5	0,5
0,5	0,5	0	0,5

Fonte: o Autor (2022).

Como resultante da média de dois lotes consecutivos (usando-se como referência os valores de cima para baixo e por coluna, conforme esquema de core), surge uma matriz

compositada contendo 8 células (Tabela 9), onde cada uma representa a média de cada lote que ingressa em intervalos de uma hora.

Tabela 9: Matriz compositada de teores colorida 2x4 (lotes 1h).

0,25	0,5	0,75	0,5
0,5	0,25	0,25	0,5

Fonte: o Autor (2022).

A próxima etapa é a comparação das médias das duas matrizes, ambas representam o material passante em uma correia transportadora durante 4h. Como esperado, o valor da média de teores não mudou, sendo igual a 0,4375%. Em contraste, as variâncias desses lotes são diferentes, também conforme esperado, pois o suporte é diferente (entenda-se aqui por suporte a massa de minério representada por cada tomada de amostra no tempo). A variância absoluta da primeira matriz é $0,0898\%^2$, e esta também é o valor da variância de dispersão. Já a variância absoluta da segunda matriz é $0,0273\%^2$, que corresponde também à variância interblocos. A diferença entre a variância de dispersão e a interblocos é igual à diferença entre as variâncias absolutas da matriz 4x4 com respeito à matriz 4x2, ou seja, $0,0625\%^2$. Isto é exatamente o que denominamos de variância de extensão e, por sua vez, é também a variância intrabloco da segunda matriz ($VAR_{ABSOLUTA\ 4X4} - VAR_{ABSOLUTA\ 4X2} = VAR_{DISP} - VAR_{INTER} = VAR_{EXT} = VAR_{INTRA\ 2X4}$).

É importante notar que, no exemplo, da primeira para a segunda matriz, as amostras foram compositadas de tal forma que o somatório de suas massas fosse igual à do lote original (matriz 4x4) e, por isso, não há nenhuma chance de enviesamento com respeito à média. É como quando ocorre um teste de heterogeneidade em que todas as porções do lote são analisadas, portanto, diferenças no CH_L ou no DH_L não interferirão no resultado final da média. Em suma, o lote foi “esgotado”.

Na prática, quando a amostra é selecionada em intervalos de tempo maiores na usina, são deixadas expressivas porções de massa do lote (na realidade, a maior parte) de fora dessa amostra. Portanto, é natural presumir que além de um erro de extensão com respeito à variância,

poderão ocorrer erros de viés significativos em torno da média. A depender do DH_L , CH_L e variabilidade espaço temporal do lote que abastece a usina, erros significativos poderão existir. Ou seja, se a coleta ocorre sempre no mesmo intervalo de tempo, e as massas dos lotes variam muito, bem como as condições de blendagem, as estimativas de variâncias de extensão podem falhar. As estimativas costumam apresentar menores erros quanto mais representativo for o teste variográfico diante das grandes massas que alimentarão a usina ao longo de sua vida útil. Todos estes fatores interferem diretamente na reconciliação, porém, são, na maior parte das vezes, ignorados.

Ainda no exemplo numérico, para fins de reconciliação, serão consideradas 3 hipóteses com respeito à situação da amostragem da matriz 2x4: (i) amostragem acurada (sem viés em torno da média) e com estimativa de erro de variância, (ii) amostragem não acurada (com viés em torno da média) e com erro significativo de estimativa de variância (quando, por exemplo, nem sequer variância de extensão é considerada) e (iii) amostragem não acurada nem com respeito à média, nem com respeito à variância, mas com estimativa dos dois erros para aferição da representatividade de fato. A partir desta última hipótese, surge uma nova proposta de QAQC para fins de reconciliação e prevenção de multas contratuais onde o foco não é apenas o valor médio do elemento em questão.

- i. **Amostragem acurada e com estimativa de erro de variância:** primeiramente, ainda que as amostras fossem preparadas evitando os erros qualitativos da amostragem em laboratório, dificilmente o valor médio de teor do contaminante MgO seria equivalente ao do lote primário de um dia de produção, pois o somatório das massas das amostras que chegam no laboratório é 4 milhões de vezes menor (Equação 58) e 20 milhões de vezes se a amostra final possuir 100 g (Equação 59).

$$\frac{2.000 \frac{t}{h} * 8 \text{ horas}}{0,5 \text{ kg} * 8 \text{ amostras}} = 4.000.000 \quad \text{Equação 58}$$

$$\frac{2.000 \frac{t}{h} * 8 \text{ horas}}{0,1 \text{ kg} * 8 \text{ amostras}} = 20.000.000 \quad \text{Equação 59}$$

Porém, é considerado que não houve viés na média das amostras em relação ao lote primário. Neste caso, ainda é preciso considerar as mudanças de

variância para propósitos de reconciliação, conforme o cálculo da variância relativa de extensão no menor intervalo da Equação 60.

$$VAR_{rel} = \frac{VAR_{EXT}}{m^2} = \frac{0,0625}{0,4375^2} = 0,326 = 33\% \quad \text{Equação 60}$$

A mina previu, nesse exemplo, um valor médio para o MgO de 0,4375%, o mesmo previsto pela usina, por outro lado, não existe estimativa da imprecisão em torno desse valor. Assim, uma incerteza só poderia ser assumida com respeito à variância calculada pela metodologia aplicada na usina (33%). Estimando para um intervalo de confiança de 95,44%, o teor de contaminante no minério poderia variar entre -0,06 e 0,94%, conforme Equação 61. A imprecisão é tão grande que o valor mínimo é menor que zero (negativo, sem sentido físico) e o valor máximo é acima do permitido (0,5%).

$$\begin{aligned} \text{Variação de teores} &= m \pm (2\sigma * m) = 0,4375 \pm (2 * 0,5714 * 0,4375) \\ &= 0,4375 \pm 0,50 \end{aligned} \quad \text{Equação 61}$$

- ii. **Amostragem não acurada e com erro significativo de estimativa de variância:** considerando uma variância causadora de viés devido ao protocolo de preparação, que inclui erros de seleção e segregação, de cerca de 5% relativo (número adotado como erro admissível entre pares de duplicatas de final de protocolo do exemplo), somada erroneamente com a variância relativa da segunda matriz 2x4 de 14,28%, teremos um total aproximado de 19,3% em torno da média. Os resultados para o mesmo intervalo de confiança de 95,44 % (Equação 62) apontam para uma variação do teor do contaminante entre 0,05 e 0,83%.

$$\begin{aligned} \text{Variação de teores} &= m \pm (2\sigma * m) \\ &= 0,4375 \pm (2 * 0,4391 * 0,4375) \\ &= 0,4375 \pm 0,38 \end{aligned} \quad \text{Equação 62}$$

Nesse caso, a interpretação das análises tem grandes chances de ser errônea por parte do profissional responsável. O uso da variância relativa do

lote no cálculo tem como resultado a estimativa de menores variações de teores em relação ao caso anterior. Assim, a variabilidade dos teores que entram na usina será diferente da prevista, por conseguinte, ocorrerão problemas na reconciliação.

iii. Amostragem não acurada com respeito à média e variância, mas com estimativa dos erros: esse caso culmina com a proposta teórica de QAQC baseada na estimativa do erro de extensão, assim possibilitando avaliar os dois tipos de erros, acurácia e precisão, conforme apresentado na equação da representatividade (Equação 21). Caso esses não sejam considerados, o lote poderá estar completamente longe do padrão de qualidade.

Em relação ao viés da média, deve ser feita a diferença do valor medido (mina) em relação ao valor de referência, em seguida, calculada a raiz do resultado $((0,5\% - 0,4375\%)^{1/2} = \pm 0,25\%)$, assim é definido o valor total de erro permitido, este deveria ser estabelecido pela usina.

Ainda nesse caso, em relação às estimativas de precisão, é recomendado realizar os cálculos para um intervalo temporal menor (não mais 1/h, mas de 30 em 30 min). E se, neste novo intervalo, os resultados ultrapassarem o valor do erro máximo permitido, o intervalo deverá diminuir novamente, podendo esbarrar em limitações que dizem respeito a outras variáveis como: mecanismo do equipamento amostrador, massas dos lotes verdadeiros e tamanhos das partículas. Nesse contexto, outro erro pode ser acrescentado, o de ponderação, porque as massas dificilmente são homogêneas no tempo. Portanto, apenas a amostragem automática não garantiria a qualidade do processo.

Então, definidas as hipóteses, fica claro que a variância de dispersão mudará conforme os intervalos de coleta variam, dependendo diretamente do DH_L e CH_L temporal do lote. **A variância de dispersão é o IHL do lote verdadeiro e sua estimativa se torna mais fiel à medida que o intervalo amostral diminui.** Caso o intervalo fosse de 0 minutos, a heterogeneidade intrínseca seria a melhor estimada possível. Na indústria, entretanto, é comum que ocorra o oposto, o intervalo de coleta aumenta à medida que a taxa de produção aumenta, pois existem limitações nos amostradores e na quantidade de análises químicas que podem ser realizadas.

Para demonstrar numericamente a problemática consequente de adotar poucas amostras para representar um lote muito maior, o exemplo com matrizes será retomado. Novamente, duas situações serão descritas: (i) as amostras são tomadas de uma em uma hora de maneira intercalada, simulando um teste de alta frequência e (ii) as amostras são tomadas de 2 em 2 horas, formando uma matriz 1x4.

- i. Amostras tomadas de 1 em 1h:** diferente da Tabela 9 em relação à Tabela 8, quando todos os valores foram compositados e utilizados na estimativa da mudança de suporte, nessa situação, foram tomadas amostras intercaladas, de cima para baixo por coluna e da esquerda para a direita na matriz primária, de modo a simular um teste de alta frequência com tomadas de uma hora em uma hora. Para melhor compreensão da situação, a matriz 4x4 foi retomada com diferente esquema de cores na Tabela 11 e a Tabela 11 representa o resultado dessas tomadas intercaladas.

Tabela 10: Matriz de teores colorida para teste de alta frequência 4x4 (lotes 30 min).

0,5	1	0,5	0,5
0	0	1	0,5
0,5	0	0,5	0,5
0,5	0,5	0	0,5

Fonte: o Autor (2022).

Tabela 11: Matriz de teores colorida 2x4 (lotes 1 h).

0,5	1	0,5	0,5
0,5	0	0,5	0,5

Fonte: o Autor (2022).

Comparando os resultados entre as matrizes 2x4, é possível perceber que acontece um aumento das variâncias de forma geral na não compositada, a variância de dispersão é maior que o dobro do caso compositado. Por

consequente, a média foi superestimada e surge uma outra variância de extensão que possui um valor menor, capaz de mascarar erros (Tabela 12). Todas essas discrepâncias decorrem da massa que deixou de ser coletada e o consequente aumento da massa a ser representada por cada tomada.

Tabela 12: Comparativo entre matrizes 2x4.

Amostragem e intervalo	Média (%)	VAR _{ABS} (% ²)	VAR _{INTRA} (% ²)	VAR _{INTER} (% ²)	VAR _{DISP} (% ²)	VAR _{DISP} REL	VAR _{EXT} (% ²)	VAR _{EXT} REL
Compositada, representado 1 em 1h	0,4375	0,0273	0,0625	0,0273	0,0898	0,4694	0,0625	0,3265
Não compositada, de 1 em 1h	0,5000	0,0625	0,1250	0,0664	0,1914	0,7656	0,0273	0,1094

Fonte: o Autor (2022).

- ii. **Amostras tomadas de 2 em 2h:** diferente da Tabela 9 em relação a Tabela 8, quando todos os valores foram compositados e utilizados na estimativa da mudança de suporte, a Tabela 13 simula uma retirada de amostras com o dobro do intervalo em relação a Tabela 9, ou seja, apenas metade da massa amostral do caso i é considerada na análise. A Tabela 13 pode ser interpretada como sendo a primeira linha da Tabela 8. Nesse caso, a média dos teores foi distorcida em comparação com as tomadas compositadas, apresentando um valor de 0,625%. A variância absoluta dos teores foi de 0,04687%², menor que o resultado para o teste de 30 em 30 minutos (0,08984%²), a diferença entre as variâncias relativas foi considerada o erro de extensão em relação ao suporte inicial, 0,3194%.

Tabela 13: Matriz de teores 1x4 (lotes 2h).

0,5	1	0,5	0,5
-----	---	-----	-----

Fonte: o Autor (2022).

As distorções nos valores de média e variância são esperadas nesse caso, pois foram utilizadas massas menores para representar um lote maior. Essa situação é semelhante ao que ocorre normalmente na indústria, onde a

quantidade de amostras não é representativa do lote inteiro. Uma variância pequena como resultado da amostragem pode influenciar o responsável pela interpretação dos dados em acreditar que a capacidade da sua blendagem de diminuir a DH_L é satisfatória. Assim, é comum que aconteçam erros nas estimativas e, novamente, problemas na reconciliação, resultado do viés nas médias e da variabilidade do minério subestimada ou até superestimada.

Dessa forma, apenas aumentar o número de incrementos não garante que a variância de dispersão seja melhor estimada por dois principais motivos: as massas das amostras permanecem muito menores que a massa do lote verdadeiro e a variância do lote muda com o tempo como resultado de fenômenos complexos (fenômeno de liberação, relação de distribuição de tamanhos, densidades, formas e teores).

Após análise crítica das diferentes situações que podem envolver testes variográficos e o erro de extensão, cabe retomar, na sequência, a fórmula de Gy em seus dois formatos, original (Equação 31) e calibrada por testes de heterogeneidade (Equação 57), para que seja possível apontar as fragilidades na união dos modelos. Então, embora executados esforços para anular todos os erros qualitativos nas tomadas de amostras, o fator de redução entre a massa do lote e a massa da amostra (M_s) sempre será grande, diminuindo minimamente apenas quando somadas as massas das amostras ao longo de toda a vida da mina. Também, a heterogeneidade de distribuição nunca poderá ser reduzida a ponto de ser desconsiderada, principalmente quando grandes massas estão envolvidas, como pilhas de blendagem, ou com variabilidade de massa no tempo, como em correias transportadoras. Ainda, por conta desses apontamentos, apenas um comparativo de duplicatas ao final do protocolo discreto já não é capaz de garantir um bom programa de QAQC que contemple os dois modelos. Por fim, é importante destacar, visto que é um fato frequentemente confundido, a fórmula do erro fundamental prevê apenas o erro mínimo causado pela IH_L , podendo haver influência de outros erros, como o próprio erro de extensão. Para sugerir um modelo único que contemple a amostragem discreta e contínua satisfatoriamente, são necessários mais testes e simulações com variações de diversos parâmetros. Dessa maneira, o modelo seria capaz de ser adaptável às mais diferentes situações (tipo de minério, parâmetro de interesse, granulometria, entre outras).

$$s^2(FSE) = \left(\frac{1}{M_s} - \frac{1}{M_L}\right)IH_L$$

Equação 31

$$\mathbf{Ln}(s_{(FE)}^2 \cdot M_S) = \alpha \cdot \mathbf{Ln}(d_N) + \mathbf{Ln}(K)$$

Equação 57

5. ESTUDO DE CASO

Como mencionado anteriormente, será apresentado, de maneira resumida, um estudo de caso onde testes variográficos personalizados possibilitaram a análise das variâncias dos erros, em especial do erro de extensão, em teores de enxofre e cinzas em minério de carvão. Através de comparações entre os resultados dos testes variográficos e da amostragem no cliente (amostragem em pilhas), vieses que comprometiam o cumprimento de cláusulas contratuais (limites máximos permitidos de 1,5% para enxofre e 55% para as cinzas) foram identificados. O estudo completo publicado por Tomazi & Souza (2021) está no ANEXO I deste trabalho.

A metodologia aplicada ao teste variográfico assemelha-se ao descrito na Seção 3.3.1. Foram analisados 3 lotes de 100 t cada com tomadas de 30 em 30 minutos (taxa de 30 t/h). Foram coletadas 50 amostras do Lote/pilha 1 (500 kg), 50 amostras do Lote/pilha 2 (500 kg) e 39 amostras do Lote/pilha 3 (500 kg): total de 1,5 t. Junto às duplicatas, totalizaram 175 amostras. Do total, 20 foram obtidas de “zero em zero minutos”, ou seja, sequencialmente em uma única parada da correia. Todas as amostras foram cuidadosamente embaladas em separado e levadas ao laboratório para homogeneização. Após esta etapa, foram peneiradas em malha de 4,75 mm. O retido foi britado em britador de mandíbulas com 4,75 mm de APF. Após nova homogeneização, foram quarteadas no Jones maior (16 células de 30 mm) e, em seguida, em um Jones menor (10 células de 8 mm). Cerca de 12 a 20 gramas foram enviadas ao laboratório. O resultado comparativo entre os testes de rotina no cliente e os encontrados no teste variográficos estão na Tabela 14.

Tabela 14: Teores de cinzas e enxofre no ROM e duplicatas: comparação entre resultados obtidos no teste variográfico e os resultados das amostragens de rotina da mineradora e do cliente.

% Cinzas ($\leq 55\%$)			% Enxofre ($\leq 1,5\%$)			
	Teste	Rotina	Dif. (%)	Teste	Rotina	Dif. (%)
Lote 1	36	38	-5,4	1,1	1,2	-8,7
Lote 2	42	35	18,2	0,8	1,6	-67,0
Lote 3	55	60	-8,7	0,9	1,3	-36,4
Média	44	44	1,4	0,9	1,4	-37,2

(*N* = número de amostras: 3 de rotina e 139 do teste)

% Cinzas ($\leq 55\%$)			% Enxofre ($\leq 1,5\%$)			
	Teste	Rotina	Dif. (%)	Teste	Rotina	Dif. (%)
Lote 1	35	37,5	-7	0,8	1,2	-40
Lote 2	43,7	45	-3	0,9	1,6	-56
Lote 3	54,2	50	8	0,7	1,3	-60
Média	44,3	44,2	-0,6	0,8	1,4	-52

(*N* = número de amostras duplicatas: 6 de rotina e 36 do teste)

Fonte: Tomazi (2021).

As médias dos teores de cinzas não diferiram significativamente, oposto do enxofre, elemento menor, para o qual já eram esperados erros maiores. As diferenças maiores foram para o enxofre nos Lote 2 e 3. Inclusive, de acordo com a amostragem do cliente, o lote 2 estaria acima do permitido (1,5%), acarretando em multa.

Em um segundo momento, os erros foram calculados a partir dos variogramas modelados pelo *SGems* e com base no banco de dados do teste variográfico com ajuda do algoritmo *gammabar* do *GSLib*. Em intervalos de tempo mínimo de coleta, 30 minutos, os erros de extensão resultaram em faixas de [0,18% a 1,54%] e [29% a 59%] para o enxofre e as cinzas respectivamente (IC = 95%), para intervalos de confiança maiores, existiria dúvida sobre a aceitação ou não do lote. Os resultados estão contidos na Tabela 15.

Tabela 15: Variância dos erros por intervalo de coleta e intervalos de teores.

Enxofre (Média = 0,85%)					Erro em % (IC=95%)	Imprecisão (%)	
Intervalo de tempo	Var ext	Var sequenc.	Var dup lab	Var total	2Desvios Rel	Lim Min	Lim Max
30/30 min	0,157			0,1660	81,48	0,157	1,543
2/2 horas	0,160	0,008556	0,000625	0,1687	82,14	0,152	1,548
4/4 horas	0,168			0,1769	84,11	0,135	1,565
Cinzas (Média = 44%)					Erro em % (IC=95%)	Imprecisão (%)	
Intervalo de tempo	Var ext	Var sequenc.	Var dup lab	Var total	2Desvios Rel	Lim Min	Lim Max
30/30 min	0,029000			0,0301	34,67	28,744	59,256
2/2 horas	0,038000	0,0009	0,000156	0,0391	39,53	26,609	61,391
4/4 horas	0,042000			0,0431	41,50	25,740	62,260

Fonte: Tomazi (2021).

Os erros de extensão resultantes do teste variográfico mostraram-se elevados mesmo considerando o menor intervalo de tempo de coleta, situação que provavelmente pioraria com um fluxo mássico maior. Recomendou-se adaptar os amostradores automáticos da mineradora para coletar amostras em intervalos menores do que estava habituada a fazer (2/2h), pois erros associados a esse intervalo não eram levados em conta. No cliente, recomendou-se buscar outra alternativa na tomada de amostras, não sendo em pilhas, e houvesse um esforço para respeitar a Lei da Equiprobabilidade. Por fim, para ambos, é importante conhecer a variância intrínseca do material/lote e as fontes dos erros associados a seu programa de garantia e controle de qualidade. O ideal seria que o teste executado fosse repetido com certa periodicidade de modo a representar o depósito, embora existam limitações de custo, tempo e capacitação necessária.

6. CONCLUSÃO

Levando em consideração o objetivo que este trabalho se propôs, exemplificar erros de extensão na amostragem de minérios, foi possível concluir que a estimativa desse erro é fundamental em um empreendimento mineiro para fins de reconciliação a curto e médio prazo e para coibir multas contratuais. A disseminação da existência do erro associado a extrapolar os valores encontrados nas amostras para o lote inteiro deve ser fortalecida na indústria, possibilitando que a problemática ganhe aceitação, relevância e, com o tempo, soluções.

Metodologias de amostragem com extrema importância, mas já defasadas, ainda são adotadas sem um olhar crítico. A TOS é problemática quando tenta aplicar conceitos que funcionam em modelos discretos para modelos contínuos. Embora com os erros de amostragem qualitativa minimizados, um modelo contínuo possui correlações e variações espaciais ou temporais que impossibilitam a simples aplicação da fórmula do erro fundamental para obtenção de uma variância mínima representativa de um lote maior. Mesmo quando a fórmula é calibrada através de testes de heterogeneidade, as massas analisadas não podem ser consideradas totalmente representativas do lote, pois o lote possui dimensões e ordens de grandezas drasticamente maiores, outro suporte.

Através de modelos teóricos e práticos, o trabalho demonstrou que, usando a variância de extensão, o erro em relação aos valores reais pode ser aproximado de forma mais adequada do que a maneira praticada atualmente na indústria. Nesse sentido, programas de QAQC deveriam garantir que os protocolos de amostragem fossem representativos, mas carecem de detalhamento e possuem como principal ferramenta a análise das duplicatas de final do protocolo. Entretanto, essas análises deixam de considerar o erro de extensão, a incompatibilidade entre modelos discreto e contínuo e a incompatibilidade de diferentes suportes amostrais.

Concluindo, a solução da problemática passa pela criação de um protocolo QAQC que contemple, além das duplicatas, um maior detalhamento técnico para tomadas de amostras e estabeleça testes variográficos rotineiros, não um único para a concepção da usina, mas frequentes para conhecimento dos erros de extensão e estabelecimento do melhor protocolo de amostragem para cada situação, minério, fase da vida útil do empreendimento ou outro objetivo específico. O sucesso desta proposta depende da integração satisfatória entre os modelos de amostragem discreta e contínua. O desenvolvimento desse modelo é a proposta de trabalho

futuro a este, se utilizaria de simulações matemáticas de amostragem, onde diversos parâmetros poderiam ser emulados, criando a possibilidade da observação e comparação de resultados em diferentes cenários. A proposta culminaria com uma estrutura básica de QAQC com aplicabilidade em qualquer situação, mas possuindo pontos determinantes que poderiam ser adaptados para cada caso específico e, finalmente, resultariam em amostragens mais acuradas e precisas.

7. REFERÊNCIAS

ARMSTRONG, M. **Basic Linear Geostatistics**. Berlin: Springer-Verlag Berlin Hiedelberg, 1950.

ABZALOV M. Z. **Drilling Techniques and Drill Holes Logging**. In: **Applied Mining Geology**. Modern Approaches in Solid Earth Sciences, vol 12. Springer, 39 p., 2016.

ABZALOV, M. Z. **Geostatistical approach to the estimation of sampling precision**. In: World Conference on Sampling And Blending (WCSB5), 5., 2011, Santiago. Santiago: GECAMIN, V.1, 243-250 p., 2011a.

ABZALOV, M. Z. **Sampling errors and control of assay data quality in exploration and mining geology**. In: Ivanov O (ed) Application and experience of quality control. InTECH, Vienna, 611–644 p., 2011b.

ABZALOV, M. Z. **Quality control of assay data: a review of procedures for measuring and monitoring precision and accuracy**. Exploration and Mining Geology, Vol.17, No 3-4,131-144 p., 2008.

CHIEREGATI, A. C. & PITARD, F. F. **Amostragem**. In: **Tratamento de Minérios**. LUZ, A.B.; FRANÇA, S.C.A. & BRAGA, P.F.A. (org.), 6a ed., CETEM, cap. 2, p. 27-65, Rio de Janeiro, 2018.

CHIEREGATI, A. C., PITARD, F. F. **The challenge of sampling gold**. World Conference On Sampling and Blending (WCSB4), 4., 2009, Cape Town. Cape Town: SAIMMV.1, 107-112 p., 2009.

CHIEREGATI, A. C., DELBONI Jr., H., COSTA, J. F. C. L., CARNEIRO, F. B. **Reconciliação pró-ativa em empreendimentos mineiros**. REM: Revista Escola de Minas, [S.l.], v. 61, n. 3, p. 297-302, 2008.

CHIEREGATI, A.C., COSTA, J.F.C.L., DELBONI Jr., H., PERONI, R.L. **Proactive reconciliation in mining industry**. Proceedings of the Third World Conference on Sampling and Blending, Porto Alegre, 176-185 p., 2007.

DE LAS CUEVAS, C. **Use of geostatistical analysis to design sampling programs in salt formations**. Acta Geologica Hispanica. 31. 67-76 p., 1996.

DEUTSCH, C.V. & JOURNEL, A. G. GSLIB. **Geostatistical Software Library And User's Guide**. New York: Oxford University Press, 1998.

ESBENSEN, K. H. & ENGSTRÖM, K. **Industrial variographic analysis for continuous sampling system validation**. Spectroscopy Europe, Weinheim, v. 29, n. 6, 14-17 p., 2017.

FRANÇOIS-BONGARÇON, D., & Gy, P. **The most common error in applying Gy's Formula in the theory of mineral sampling, and the history of the liberation factor**. Journal of the South African Institute of Mining and Metallurgy, 475-479 p., 2002.

FRANÇOIS-BONGARÇON, D. Error variance information from paired data: applications to sampling theory. **Exploration and Mining Geology**, Montreal, v.7, n.1-2, 161-165 p., 1998.

GEELHOED, B., **Is Gy's formula for the Fundamental Sampling Error accurate? Experimental evidence**, Minerals Engineering, vol. 24, 169-173 p., 2011.

GOES, M.A.C., POSSA, M. V., LUZ, A.B., **Amostragem de Minérios**, Centro de Tecnologia Mineral, Departamento de Estudos e Desenvolvimento, Rio de Janeiro, RJ, Brasil, p. 5-24, 1991.

GOOVAERTS, P., **Geostatistics for Natural Resources Evaluation**, Department of Civil and Environmental Engineering, The University of Michigan, Ann Arbor, Oxford University Press, New York – NY, USA, p. 89, 1997.

GUEDES, T.A. & MARTINS, A.B.T., **Estatística Descritiva**, Projeto Aprendendo Fazendo Estatística, Universidade Estadual de Maringá, Maringá - PR, Brasil, p. 1 - 35, 2016.

GY, P. M. **Sampling of discrete materials** — a new introduction to the theory of sampling - I. Qualitative approach, Chemometrics and Intelligent Laboratory Systems, edição 74, Cannes, France, p. 7-24, 2004a.

GY, P. M. **Sampling of discrete materials** - II. Quantitative approach—sampling of zero-dimensional objects, Chemometrics and Intelligent Laboratory Systems, edição 74, Cannes, France, p. 25-38, 2004b.

GY, P. M. **Sampling of discrete materials** - III. Quantitative approach—sampling of one-dimensional objects, *Chemometrics and Intelligent Laboratory Systems*, edição 74, Cannes, France, p. 25-38, 2004c.

GY, P. M. **Sampling for analytical purposes**. GY, P. *Sampling for analytical purposes*. West Sussex, England: John Wiley & Sons, 1998.

GY, P. M. **Sampling of heterogeneous and dynamic material systems**. In *Theory of heterogeneity, sampling and homogenizing*, Amsterdam: Elsevier, 1992.

GY, P. M. **Sampling of particulate materials** – Theory and Practice. Amsterdam: Elsevier, 1982.

GY, P. M. **Sampling of particulate materials** – Theory and Practice. Amsterdam: Elsevier, 1979.

HUMPHREYS M., SHRIVASTAVA P. **Choosing an exploration drillhole spacing: a case study in an iron mine**. *Geostatistics Wollongong 96 - Proceedings of the Fifth International Geostatistics Congress*, Wollongong, Australia, 780-791 p., 1996.

ISAAKS, E.H. & SRIVASTAVA, M.R., **An Introduction to Applied Geostatistics**, Oxford University Press, New York, 3-382 p., 1989.

JOURNEL, A. G. e HUIJBREGTS, C. J. **Mining Geostatistics**. New York: Academic Press, 1978.

KOPPE, V.C. & COSTA, J.F.C.L. & KOPPE J.C., **Coordenadas cartesianas x coordenadas geológicas em geoestatística**: aplicação à variável vagarosidade obtida por perfilagem acústica, *Revista Escola de Minas*, Ouro Preto – MG, Brasil, p.26, jan. 2006.

LARSON, R. & FARBER B., **Estatística Aplicada**, 4ª edição, Editora Pearson, São Paulo – SP, Brasil, p. 31-100, 2010.

LUIZ, T. S. J. **Uso de variância de extensão na amostragem dos parâmetros de águas minerais**. Tese apresentada ao programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, UFRGS, Porto Alegre, RS, 2021.

LUIZ, T. S. J. & SOUZA, V. C. G. & KOPPE, J. C. **Use of extension variance in monitoring of fluoride in bottled water**. *Water Supply*. 20. 3281-3287. 10.2166/ws.2020.232, 2020.

MARQUES, D. M. & COSTA, J. F. C. L. **Choosing a proper sampling interval for the ore feeding a processing plant: a geostatistical solution.** International Journal of Mineral Processing, Amsterdam, v. 131, p. 31-42, 2014a.

MARQUES, D. M. & COSTA, J. F. C. L. **Analysis of the dispersion variance using geostatistical simulation and blending piles.** Journal of the South African Institute of Mining and Metallurgy, Amsterdam, v. 114, p. 587-599, 2014b.

MARQUES, D. M. **Métodos estocásticos aplicados a definição de estratégias de amostragem e homogeneização.** Tese apresentada ao programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, UFRGS, Porto Alegre, RS, 2014c.

MARQUES, D. M. & COSTA, J. F. C. L. **An algorithm to simulate ore grade variability in blending and homogenization piles.** International Journal of Mineral Processing, Amsterdam, v. 120, 48-55 p., 2013.

MATHERON, G. **The theory of regionalized variables and its applications.** Fontainebleau: École Nationale Supérieure des Mines de Paris, 1971.

MATHERON, G. **The selectivity of the distributions and "the second principle of geostatistics".** In: Geostatistics for natural resources characterization, Lake Tahoe, USA, Reidel Publishing Company, p. 421-433, 1984.

MINNITT, R., RICE, P. M., & SPANGRNBRTG, C. Part 2: **Experimental calibration of sampling parameters K and alpha for Gy's formula by the sampling tree method.** Journal of the South African Institute of Mining and Metallurgy, 107, 513–518 p., 2007.

MONTGOMERY, D.C.; RUNGER, G.C. **Applied Statistics and Probability for Engineers.** Phoenix: John Wiley & Sons, 768 p., 2010.

MORAES, F. P.; SILVA, R.C.F.; ROSIÉRE, C. A.; DIAS, G. S. **Caracterização dos contaminantes do minério de ferro do depósito mineral da Serra do Sapo, Conceição do Mato Dentro, Minas Gerais,** Geologia, Revista do Instituto de Geociências da Universidade de São Paulo, São Paulo, SP, Brasil, p. 81-100, 2020.

MORETTIN, A. M & BUSSAB, W.O., **Estatística Básica**, 5ª edição, Editora Saraiva, São Paulo, SP, Brasil, 2004.

MUCHA, J. e SZUWARZYŃSKI, M. **Sampling errors and their influence on accuracy of zinc and lead content evaluation in ore from the Trzebionka mine** (Silesian–Cracow Zn–Pb ore district, Poland). *Chemometrics and Intelligent Laboratory Systems*, Amsterdam, v.74, n.1, 165-170 p., 2004.

NAPIER-MUNN T.J., MORRELL S., MORRISON R.D. **Mineral Comminution Circuits. Their operation and optimization**, JKRM, 1996.

PYTHONIC EXCURSIONS, **Stories About Python and Data Science**, disponível em: <https://aegis4048.github.io/jupyter_images/search_template_1.png>, acesso em 25/04/2020.

PITARD, F. F. Pierre Gy's **Sampling Theory and Sampling Practice**. 2nd edition. New York: CRC Press, 488 p., 1993.

PITARD, F. **Exploration of the “Nugget Effect”**. In: *Geostatistics for the Next Century. Quantitative Geology and Geostatistics*, v.6 (Dimitrakopoulos, R. Editor), Springer, Dordrecht, 1994.

ROYLE, A.G., NEWTON, M.J., SARIN, V.K. **Geostatistical Factors in Design of Mine Sampling Programmes**. *Transactions of the Institution of Mining and Metallurgy, Section A: Mining Technology*, 81 (785), a82-a88 p., 1972.

SINCLAIR A. **Geostatistics and The Earth Sciences**. *West Miner. Volume 51*. 24-26 p., 1978.

SINCLAIR, A. J. & BLACKWELL, G. H., **Applied Mineral Inventory Estimation**, Cambridge University, UK, 81 p., 2004.

SOUZA, V. C. G., COSTA, J. F. C. L., SOUZA M. L., **Curso de Introdução à teoria de amostragem de Gy**, MMD00143, Material de apoio da Disciplina, Programa de Pós Graduação em Engenharia de Minas, UFRGS, Porto Alegre, RS, 2021.

TOMAZI, G. L., SOUZA, V. C. G., **The most ignored error in the history of mining sampling – extension variance: a case study applied to coal**, *International Journal of Coal Preparation and Utilization*, 2021.

YAMAMOTO, J. K. & LANDIM, P. M. B., **Geoestatística - Conceitos e Aplicações**. 1ª edição, Editora Oficina dos Textos, São Paulo - SP, Brasil, p. 19-141, 2013.

WACKERNAGEL H. **Extension and Dispersion Variance**. In: Multivariate Geostatistics. Springer, Berlin, Heidelberg, 1998.

ZANON, J. (2016). **Determinismo e Estocasticidade - Histórico e Ecologia**. Universidade de Goiás, Programa de Pós-Graduação em Ecologia e Evolução. Brasil, Goiânia/GO, p. 13. 2016.

ANEXO I



The most ignored error in the history of mining sampling – extension variance: a case study applied to coal

Gustavo Leandro Tomazi  and Vládia C. G. De Souza

Department of Mining Engineering, Federal University of Rio Grande do Sul

ABSTRACT

Apart from the sampling errors related to the sample preparation protocol in a laboratory to determine the ash and sulfur content of the coal, the delimitation/extraction errors (recovery, among other geological factors) and the "extension" error must be considered. As the collection time interval increases, the extension error also increases, so that even automatic sampling has an error that is often completely ignored. The extension error was determined via a variographic test, which was specifically designed for this study. This test was proposed to be carried out at the mining company to check the uncertainties (bias and reproducibility errors), regarding the sampling at the client. In the latter, sampling is done in piles with manual samplers. At the mining company, there are automatic collectors, but the sample interval is not considered when calculating the total error. To redeem reconciliation problems between the mining company and the client, as well as to avoid contractual fines, it was decided to compare the results of the variographic analysis carried out in the mining company with those obtained by the client. These showed considerable bias and extension errors according to the collection time intervals in the mining company, especially regarding sulfur. In shorter time intervals, 30 minutes, extension errors have resulted in a range of [0.18% to 1.52%] and [29% to 59%] for sulfur and ashes respectively (CI = 95%). For longer intervals, there would be a doubt about accepting or not the lot: maximum limits are 1.5% for sulfur and 55% for ashes. Also, as extension errors are elevated and possibly higher for "true mass flows," which exceed 2,000 t/h, it was recommended to adapt the automatic samplers of the mining company at shorter intervals than it was used to do (2/2 hours). Then, it will improve the processes control, including blending, avoiding reconciliation problems, and subsequent contractual fines. The client was alerted to the problem of bias which occurs when the coal is sampled manually in piles. For both, it is important to know the intrinsic variance of the material/lot and the errors related to the sampling.

ARTICLE HISTORY

Received 28 September 2021
Accepted 17 November 2021

KEYWORDS

Sampling error; coal; ashes; variography; heterogeneity; reproducibility

Introduction

This study aims to show the importance of the extension error: the error made regarding all lot mass that is not collected during automatic sampling on belt conveyors. The longer the collection time interval is, the greater is this error. That knowledge is essential in proactive reconciliation processes (Chierigati, et al., 2008) and helps to design the circuits and equipment of an ore treatment plant (Pitard, 1993).

CONTACT Gustavo Leandro Tomazi UFRGS  gustavo.tomazi@hotmail.com  9500, Bento Gonçalves Ave.,- sector IV, building 43426, room 217., Porto Alegre/RS, Brasil

© 2021 Taylor & Francis Group, LLC

Regarding the sampling errors, Gy established (1979, 1982, 1992) this list below, which, generally, the major is the extension variance and is often totally ignored:

- **Fundamental Error** (variation in the constitution of the material): this error would be zero in a hypothetical situation that all elements selected in the sample were independent of each other and had the same probability of being collected;
- **Segregation and Grouping Error** (variation in the distribution of material's groupings): this error is a consequence of the segregation. Because we work in the gravitational field, two elements can never be assumed to be independent. It is possible to minimize it through a good mixture with an equiprobabilistic method of collection;
- **Non-periodic error**: the error takes into account macroscopic fluctuations, or trends, observed in the general lot scale, large-scale segregation. That is, this error does not have a well-defined fluctuation. The variogram does not have a constant or stationary global variance;
- **Periodic Error** (process variation; long observation period): periodic error is related to cyclical and long-range variability in a continuous process. This can be added when there is no blending pile before feeding the industrial plant, and two materials with different heterogeneity, one with high grade and the other low grade, are constantly processed. The variogram does not have a constant or stationary global variance, and we see a cyclical pattern for long range;
- **Delimitation Error** (of increment): When taking an increment, a correct selection is reached when each belt strip is collected at the same instant of time (cutter in a straight line and with constant speed);
- **Extraction** (increment) **error**: error associated with sampling installation, operation, and maintenance of samplers. For the extraction error not to occur, it is assumed that the elements in the increment model are correctly delimited;
- **Preparation Error**: it is the sum of five independent errors, arising from the operations of particle size reduction, homogenization, and quartering to which the primary sample is subjected (loss of particles + sample contamination + unintentional change in the characteristic of interest to be measured in the sample + unintentional operator errors + intentional alteration of the characteristic of interest to be measured in the sample).

The periodic and non-periodic errors constitute a continuous selection error model proposed to estimate long-range errors of sampling. The extension error is a third far-reaching error also called an integration or interpolation error (Pitard, 1993).

Also, when sample preparation protocol errors are determined, along with analytical error; these uncertainties do not feed back into reconciliation models (Abzalov 2011, 2016; Chierigati, et al., 2008; Chierigati, 2009; Marques and Costa, 2014a, Marques and Costa, 2014b). Everyone needs to be aware that automated sampling itself does not guarantee quality control (Abzalov 2016; Castilho, 2005; Mucha and Szuwarzyński, 2004a).

We performed variographic analysis on three lots of coal from a mining company in order to show the impacts of these sampling errors (Gy 1979, 1982, 1992; Pitard, 1993; Abzalov, 2016). The methodology was based on the classical statistics (Montgomery, 2010) and geostatistics (Journel, 1978; Matheron 1971, 1984), using SGems and GSLib software (Deutsch and Journel, 1998). These lots were tracked to form three separate piles in the

customer, where the sampling protocol is run almost entirely manually: the piles are mixed with the help of dump trucks and backhoes, being flattened, and quartered *in loco*, resulting in piles of increasingly smaller masses. Our goal was to identify the sources of contractual fines and the uncertainties inherent in this mining/client reconciliation process, aiming not only to determine biases but also the variances of the ash and sulfur contents (Gy 1979, 1982, 1992; Pitard, 1993; François-Bongarçon 1998).

The three batches/piles reconstituted with their original mass were separately fed into a silo which discharged the material onto a conveyor belt. The samples were taken in the shortest time possible, to “deplete” the lot, according to the methodology to be described in detail below.

This test allowed that parts of the lot, which were not accessible in the piles, were collected. Thus, the variances as a function of time intervals could be measured. This is the variographic analysis, which allows knowing the internal heterogeneity of the lots and the resulting sampling errors.

Methodology

Mathematical Fundamentals

The extension variance refers to the precision error made when decreasing the number of increments and extrapolating the average of the increment contents analyzed for the whole population (in this case, the production of that period analyzed, denominated “batch”). This occurs analogously to the process of extrapolating the value of the grade of a core sample to an entire block of ore to be mined (Armstrong, 1950; Wackernagel, 1998; Abzalov 2011). In addition, when we extrapolate the average value of the increments collected on a conveyor belt to the entire mass flow which passed through that belt in the shift, day or month, and so on, we are making an extension error. Through variography (Esbensen and Engström, 2017), dispersion, and extension variance (Wackernagel, 1998), we identified nonrandom variances, as well as cycles in processes during the sampling. Also, through dispersion and extension variance analysis, we can ascertain precision errors (Wackernagel, 1998). This method makes it possible to ascertain errors even if the sample masses are much smaller than ten times the lot mass (limit imposed by the equations of Gy 1979, 1982, 1992; Pitard, 1993).

The variogram function is an average measure of the variance of a parameter of interest from samples or points separated by a time “h,” taken at different time intervals: h, 2 h, 3 h ... Fig. 1 shows an example of taking samples in a conveyor belt (One Dimensional Lot).

These time intervals are called lags or steps (h). Variances are calculated for the respective sets of samples with lag equal to 1 (smallest collection time interval), then lag equal to 2 (double time interval), and so on. “It would be expected” that at lag equal to 0 (zero), the variance would be zero, but this does not happen: this variance (called the nugget effect) concerns the intrinsic or internal heterogeneity of the material/lot and often is “inflated” due to sampling errors (Pitard, 1993, 1994). Among many errors, those of extraction and delimitation (when taking the sample) are included, those resulting from the choice of sampling support (number of samples, sampling interval, masses, etc.), those of preparation (crushing, homogenization, slicing), and chemical analysis respectively in order of magnitude of error (Pitard, 1993; Gy 1979, 1982, 1992; Esbensen, 2017; Esbensen

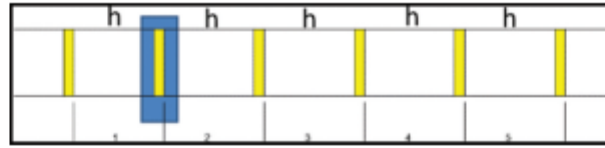


Figure 1. Sampling on conveyor belt.

and Engström, 2017). In Equation 1, the content of the “ a_i ” sample is “ j minutes apart” from the “ a_{i+j} ” content sample, where n is the number of samples and $n-j$ is equal to the number of lags or steps.

$$V(j) = \frac{1}{2(n-j)} \sum_{i=1}^{n-j} (a_{i+j} - a_i)^2 \quad (1)$$

In Fig. 2, there are the essential parameters of a semivariogram (Journel and Huijbregts 1978; Matheron 1971, 1984): the “range”; the “lags” or steps (collection time intervals); the “sill” or global variance and, finally, the “nugget effect” (a priori variance).

The variance value of the contents of the set of samples, from which the variogram stabilizes, shows there is no longer correlation of the data: this “plato” is called “sill.” The value in minutes at which this occurs is called the range. Thus, the “sill” is a constant value that takes the variogram in time intervals longer than the range.

In the case of taking samples on conveyor belts, over time, the extension variance refers to the error made when extrapolating the content of samples taken, for example, every 8 minutes for the “sample influence zone”: 16 in 16 minutes (Figure 3).

The extension variance can be considered as a long-range or large-scale error (Pitard, 1993, 1994). By assuming the sample represents a larger belt range, which is the zone of influence, a quantifiable error is made through chronostatistics, more specifically variography. This error is called by some authors as an interpolation error (Pitard, 1993, Esbensen and Engström, 2017b; Mucha and Szuwarzyński, 2004a).

The variability of the zone of influence (“actual variability”) tends to be greater than any smaller intervals which might be considered: therefore, at larger intervals, the inaccuracy may appear to be “small,” even though it is acceptable when it may not be (Matheron 1971; 1984; Journel and Huijbregts 1978; Mucha and Szuwarzyński, 2004a; Marques and Costa, 2014a; Marques and Costa, 2014b; Wackernagel, 1998). This is illustrated in Fig. 4, where sections in yellow, red, or blue represent bands of material with very different contents of the gray part.

The equation used to calculate the extension variance is (Wackernagel, 1998):

$$\sigma_{ext}^2\left(\frac{\cdot}{T}\right) = 2\bar{\gamma}\left(\frac{\cdot}{t}\right) - \bar{\gamma}\left(\frac{\cdot}{T}\right) \quad (2)$$

In this equation, the extension variance is $\sigma_{ext}^2\left(\frac{\cdot}{T}\right)$; the mean dispersion variance, concerning the smallest time interval “ t ,” é $\bar{\gamma}\left(\frac{\cdot}{t}\right)$; while, $\bar{\gamma}\left(\frac{\cdot}{T}\right)$ is the average variance of dispersion, regarding the largest time interval “ T ” (Wackernagel, 1998).

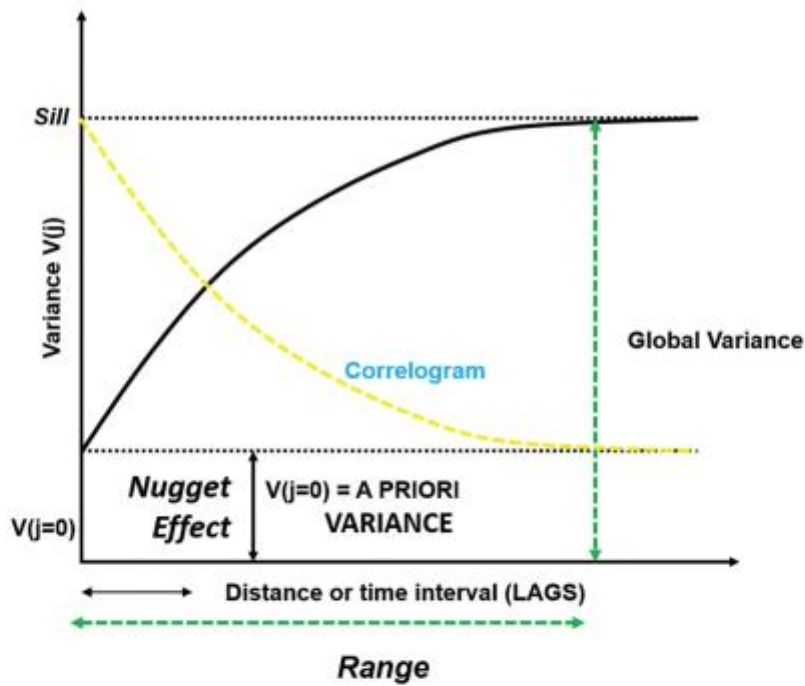


Figure 2. Variogram example.

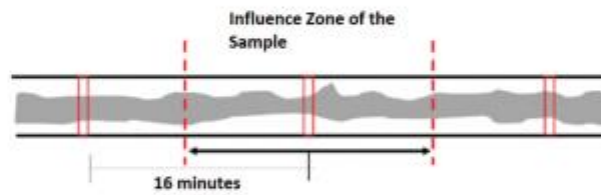


Figure 3. Extension error on the conveyor belt.

“T” is the time interval in which the extension variance is to be estimated, any interval longer than the interval “t.” The “T” time interval is twice the previous time interval “t” in Equation 2. The shortest time interval in which samples can be collected is represented by a “.” (“dot”) in this equation, in this example: every 4 minutes. If “T” is every 16 min, then “t” will be 8 min. These variances can be calculated with the GsLib algorithm called “*gammabar*.”

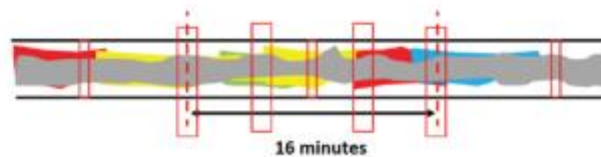


Figure 4. Example of interpolation errors.

The larger the collected masses are, the precision error tends to be smaller, but there is a physical limitation in terms of maximum capacity (collector sizing), laboratory processing (sample preparation), and costs related to the preparation and chemical analysis itself.

In general, a single “variographic test” is carried out as soon as the plant starts operating for the first time (Pitard, 1993; Esbensen and Engström, 2017; Esbensen and Wagner, 2017a; Marques and Costa, 2014a, Marques and Costa, 2014b). This consists of collecting samples through automated processes/equipment in as small-time intervals as possible. Also, this test can be performed by stopping the belt and collecting the samples, although this results in much more work.

Furthermore, the software allows to model and adjust a mathematical function capable of describing the variability of elements over time. The nugget effect (Fig. 2) can be chosen as the variance of the pairs of duplicates which must exist in a test of this type (Abzalov 2011, 2016). For example, at least one out of every 20 samples collected on the belt must be prepared to provide two subsamples (1 pair). The total of analyzed pairs gives the minimum variance of the processed material, in other words, the nugget effect or variance at an “almost zero” time interval.

Equation 3 is used to estimate the relative variance of the pairs (to compare sampling errors and do sums). Equation 4 is applied to compare the variability between pairs or different parameters (coefficient of variation).

$$\sigma^2\text{RSV} = \frac{2}{N} \sum_{i=1}^N \frac{[Z(X_i) - Z(X_i+0)]^2}{[Z(X_i) + Z(X_i+0)]^2} = \frac{2}{N} \sum_{i=1}^N \left(\frac{(a_i - b_i)^2}{(a_i + b_i)^2} \right) \quad (3)$$

$$\text{CV}\% = 100\% \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{\sigma_i^2}{m_i^2} \right)} = 100\% \sqrt{\frac{2}{N} \sum_{i=1}^N \left(\frac{(a_i - b_i)^2}{(a_i + b_i)^2} \right)} \quad (4)$$

Again, the relative variance of the sample pairs would give the value of the variance in $h \rightarrow 0$ minutes: the nugget effect. Furthermore, if this value varies from 20 to 30% of the total sample variance, it can be considered significant. Thus, the higher the contribution of the sampling process to the variability of the processed lot and the more “structured” the variogram (inadequate blending), the higher will be the uncertainty regarding lot acceptance or rejection, based on parameters provided for in the contract.

In this sense, methodologies that use geostatistical tools may be more suitable than those that use the Gy sampling fundamental error equation, which assumes that all other errors, such as segregation and clustering, have been minimized. This equation uses sampling “constants” or needs to be calibrated by some kind of “heterogeneity test.” Heterogeneity tests are often arduous; involving a long preparation time and a large number of analyzes (Minnitti et al., 2011). For Gy, each step of the sample preparation protocol needs to be calculated separately. Also, when the mass of the lot is 10 times higher than that of the sample, there is an underestimation of sampling errors, especially because the spatial/temporal correlation of each deposit (geological factors) is not being considered.

In the case of coals, it is important to know the uncertainty of the ash contents. As these are computed as oxides, it would be very complex to directly calculate the mineral contents by stoichiometry. In addition, it would not be possible a priori to extrapolate the Gy formula, replacing “mineral content” for “ash content.”

Variographic Test

This test aimed to verify differences in the measurement of the average contents of the main elements of the coals and their respective variances through sampling in piles and conveyor belts (variographic test); thus, making it possible to define the best sampling interval for automatic collectors. The biggest uncertainties reside in the sampling of the elements of lower concentration; therefore, we chose to determine also sulfur content. The studied ore originates from a mine located in southern Brazil. Coal from this region typically has high ash and sulfur content when compared to other regions of the world. Customer sampling errors impact especially the variance of this parameter.

At the client, the material received from the mining companies is dumped by trucks, one or two piles at a time, in a specific local. These undergo a homogenization by a backhoe and a division in the cone and quarter method: resulting in ever smaller piles: from 500 kg to about 5 kg. In the end, one of the four slices is selected with a small shovel, consisting of the sample. Every twenty times a shovel is removed from another slice as a duplicate.

Thus, for the variographic test, three piles were chosen, called Lots, of approximately 100 tons each. The stacks fed, one at a time, a silo. The silo material was unloaded onto a conveyor belt at an approximate rate of 30 t/h. This was stopped every 30 minutes to collect samples with a template. Within this procedure, samples were collected in duplicates (25% insertion rate, **five times higher** than the conventional one), in addition to blanks and standards, inserted during sample preparation (Abzalov 2016). These procedures are part of the sampling assurance and control program (QAQC). The template consisted of two metal plates to delimit about 2 meters of a small 40-centimeter-wide belt. The material collected between the plates corresponded to about 11 kg of the sample (wet base). These samples were carefully packaged separately and taken to the laboratory for homogenization. After this step, they were sieved in a 4.75 mm mesh. The retained material was crushed into jaws with 4.75 mm of APF (closed position opening of the crusher). After further homogenization, they were split into the larger Jones (16 cells of 30 mm) and then into a smaller Jones (10 cells of 8 mm). About 12 to 20 grams were sent to the laboratory.

In total, 50 samples from batch/pile 1 (500 kg), 50 samples from Batch/pile 2 (500 kg), and 39 samples from Batch/pile 3 (500 kg) were collected: summing 1.5 t total. 139 samples were collected and, counting the duplicates (plus 36), this results in 175 samples. There were 15 duplicates of pulp and 21 duplicates of gravel. Of the 139 samples, 20 were obtained from “zero in zero minutes,” in other words, sequentially in a single stop of the belt, during the twelfth sampling of lot 1 and the eighteenth sampling of lot 2. Lot 3 did not have sequential sampling and suffered the loss of a sample. The so-called “sequential” samples simulated the “zero” minute collection time condition. The other samples were all taken approximately every 30 minutes. Furthermore, 18 whites (“quartz PM”) and 18 standards were inserted. This test took about seven days to be performed in terms of sample collection. It took about 30 days to prepare and analyze all samples. Another 30 days were used to confirm the results, repeating the preparation and analysis process from the beginning. About 20 days were used to proceed with the resampling of the lots.

Customer Sampling Analysis

The uncertainties associated with the sampling and blending process of the studied location arise, mainly, because of the mass of lots (piles) being higher than those collected.

As the increments have been removed with a small shovel, was clearly seen that there were pieces of material larger than 15 cm left behind. Logically, these, which were not few, had no chance of being collected, increasing the chances of a bias occurring in the end. Also, as mentioned earlier, Gy (1992) will only give us the variance of the fundamental error if the probabilities of biases are completely minimized. The most worrying, at first, would be related to the biases that can occur because of the sampling in piles (Pitard, 1993; Gregorieff, 2002).

During the preparation in the laboratory, it was suggested that more care had to be taken with sample weighing and quartering. The splitters were rusty, and this could interfere with the sulfur analysis results.

QAQC Considerations

The relative variance between pairs of samples or duplicates (Abzalov 2011), according to Equation 3, will provide a good hypothesis of the nugget effect to model the variogram. From the adjustment of the mathematical function to the experimental data and the main parameters of the variogram, it is possible to proceed to the step of calculating the dispersion variances as a function of the time interval by the “gammabar” algorithm.

As for blending, the better the material is mixed, the smaller the variance in $h \rightarrow \infty$ minutes, and the nugget effect starts to contribute 100% of the global variance of the data. Thus, there are reproducibility errors due to the collection, the preparation protocol, the pre-mixing of the lot, and, finally, the extension.

In this case, the contract, which determines the acceptance of the lot and application of fines, predicts the value of ash and sulfur cannot exceed 55% and 1.5% respectively. Otherwise, there would be a contractual penalty. Also, it was agreed that the relative differences in terms of deviations, for a 95% confidence interval, could not be higher than 5% between pairs of laboratory duplicates (from the end of the preparation protocol). For duplicates of the variographic test (sequential duplicates), it could reach up to 20%.

Results and Discussion

Figure 5 shows the variation of grades overtime on the belt. Sulfur showed a pronounced variation from lot 1 to lot 3, with the highest contents in lot 1. Ash showed the highest contents at the end. Still, in Fig. 5, we have the color scale for sulfur and ash: we can visualize this over time.

Table 1 shows a comparative summary of the mean grades: variographic test versus standard pile sampling procedure. The difference (%) was calculated as the difference between the content measured by variography and that obtained from pile sampling. This result is finally divided by the content provided by the test and multiplied by 100.

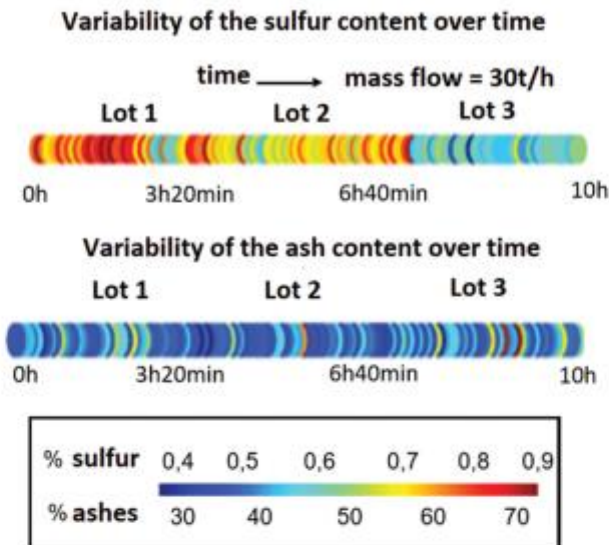


Figure 5. Timeplot on conveyor belt.

We observed that the averages of ash contents did not differ significantly, opposite to sulfur, a lower concentration element, as expected. The biggest differences were for sulfur in Lot 2 and 3. For so, according to the manual sampling, lot 2 would be above the permitted level (1.5%), which should result in a fine.

We can see in Table 1, that the client may be overestimating the sulfur contents due to the manual collection methodology.

Thus, part of the total sampling error, which is due to the intrinsic heterogeneity of the lots (CHL) and distribution (DHL) may have been minimized through the variographic analysis. We can see in the table that the reproducibility values (duplicates) are good in the case of pile sampling, especially for sulfur. Nevertheless, these numbers do not express the true error regarding all the mass of the pile.

While ash can be better distributed in the lot, the same does not occur with pyrite, causing a higher distribution heterogeneity error (DHL).

Table 1. Ash and Sulfur in ROM and duplicates: mining company x customer.

% Ash ($\leq 55\%$) % Sulfur ($\leq 1,5\%$)						
	Test	Rotine	Dif. (%)	Test	Rotine	Dif. (%)
Lot 1	36	38	-5,4	1.1	1.2	-8.7
Lot 2	42	35	18,2	0.8	1.6	-67.0
Lot 3	55	60	-8,7	0.9	1.3	-36.4
Average	44	44	1.4	0.9	1.4	-37.2
<i>(N = number of samples: 3 routine and 139 test)</i>						
% Ash ($\leq 55\%$) % Sulfur ($\leq 1,5\%$)						
	Test	Rotine	Dif. (%)	Test	Rotine	Dif. (%)
Lot 1	35	37.5	-7	0.8	1.2	-40
Lot 2	43.7	45	-3	0.9	1.6	-56
Lot 3	54.2	50	8	0.7	1.3	-60
Average	44.3	44.2	-0.6	0.8	1.4	-52
<i>(N = number of duplicate samples: 6 routine and 36 test)</i>						

Although not all lots exceeded the maximum allowed sulfur value by the test, there were significant reproducibility errors and bias. Therefore, it is previously concluded that, especially for the minor elements, the variographic analysis must be carried out to scale or adjust the automatic sampling.

Figures 6 and 7 show, respectively, the semi-variogram for sulfur and the ashes. It is noted that for sulfur there is a considerable “starting error” (high nugget effect – Fig. 6). This is evidenced by the high value of the variance of sequential samples, which was for sulfur and ash, respectively, 18% and 6%. The semi-variogram of sulfur is “unstructured.”

This extension error shown in Fig. 7 must be added to that of the duplicates (sample preparation protocol error).

The relative variance of the test duplicate pairs (**sequential**), calculated according to Equation 3, resulted in **0.0086 for sulfur**. In the graph in Fig. 7, we observe a value of extension relative **variance of 0.157** (dimensionless) for the minimum sampling interval: **30 minutes**.

When we add: $0.157 + 0.0086 + 0.000625 = 0.1660$, we will have the value of the total relative variance of the sample in this time interval. This last added value corresponds to the error of the laboratory duplicates.

To obtain the range of values of sulfur contents, we must extract the root of the value **0.1660** and multiply by 2, and so we obtain the total relative error in terms of deviation for 95% CI. This error was $\pm 81.5\%$; considering this, we will have the following uncertainty: $0.85\% (\text{mean}) \pm (0.85\% \times 0.815) = 0.85\% \pm 0.69\%$. The sulfur content of the lots, 95 out of 100 times, can then range from **0.157% to 1.543%**. As the maximum allowed would be 1.5%, this error would be significant. In the case of customer sampling, we are not even aware of this error. For lots with average sulfur values equal to 1%, the range is: **[0.185% to 1.815%]**. For confidence intervals of 99.9% and mean equal to or greater than 0.85%, the minimum sulfur contents assume negative values, which does not make physical sense, denoting the high magnitude of uncertainties involved in this sampling process.

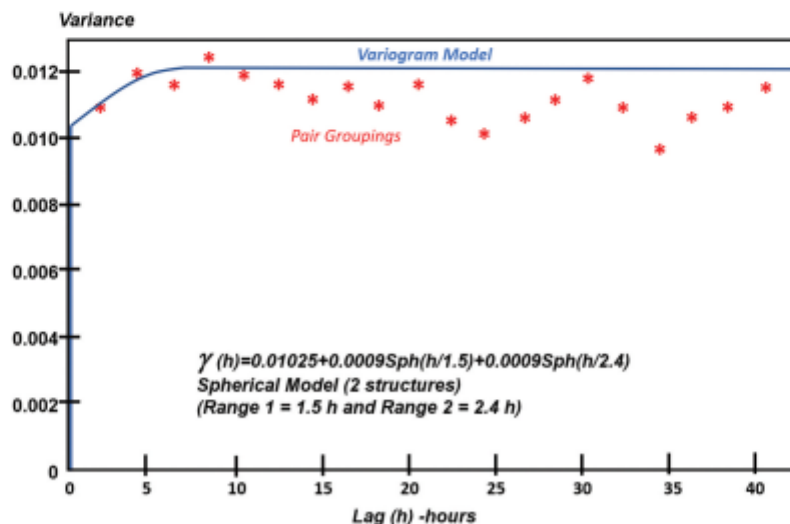


Figure 6. Sulfur semi-variogram.

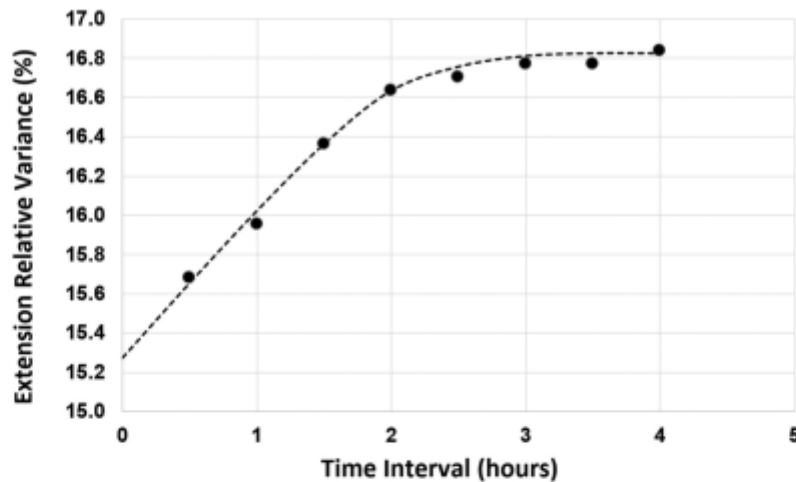


Figure 7. Extension error for sulfur.

For the larger sampling interval, such as every 4 hours, the **relative** variance reaches **0.168**, as seen in Fig. 7. This value multiplied by the mean squared is the total absolute variance of the data set) that appears next to the variogram plateau in Fig. 6.

Figure 8 shows the ash semi-variogram, and Fig. 9 shows the extension errors over time intervals.

The relative variance of the test duplicate pairs (**sequential**) for **ash** resulted in **0.0009**. In the graphic in Fig. 9, we observe an extension relative **variance** value of **0.029** (dimensionless) for the minimum sampling interval: **30 minutes**. By adding: $0.029 + 0.0009 + 0.000156 = 0.03$, we will have the value of the total relative variance of the sampling in this time interval. This error was $\pm 34.7\%$; considering this, we will have the following uncertainty: $44\% \text{ (mean)} \pm (44\% \times 0.347) = 44 \pm 16.6\%$. The ash content of the lots, 95 out of 100 times, can then range from **28.7% to 59.2%**. As the maximum allowed would be **55%**, this could also be considered a significant error.

When sampled from 4/4 hours this range for the same CI is increased to: **25.7% to 62.3%**. Extension errors are higher when calculated for larger amounts of mass ("true" mass flows). In the mining company passes more than 2,000 t/h for example. A summary of errors for different collection times can be seen in Table 2.

Conclusion

The extension errors resulting from the applied variographic analysis proved to be high, even considering the shortest collection time interval (30/30 min), especially for sulfur. Possibly, these will be higher when calculated for the "true" mass flows, which exceed 30 t/h (mass flow of the test).

Thus, it was recommended to adapt the automatic samplers in the mine in order to collect samples at shorter intervals than it was used to (2/2 h). In this way, the company will be able to improve the control of its processes, including blending, avoiding reconciliation problems, and subsequent contractual fines. The client was alerted to the problem of bias

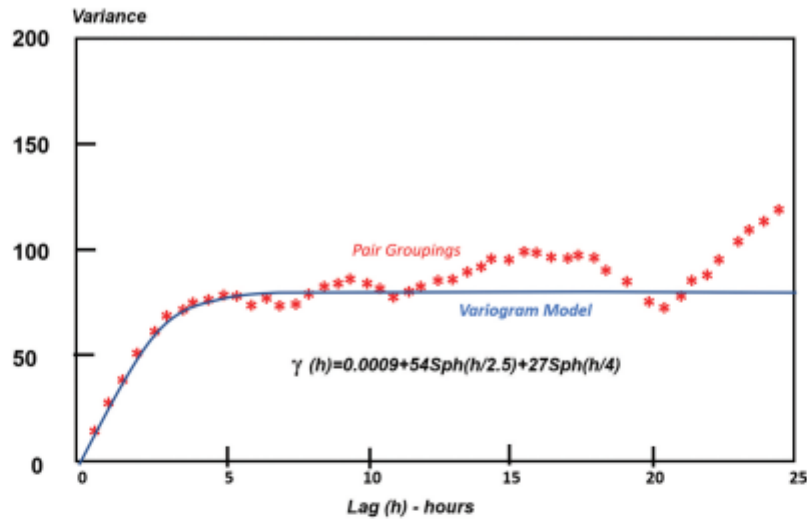


Figure 8. Semi variogram for ash content.

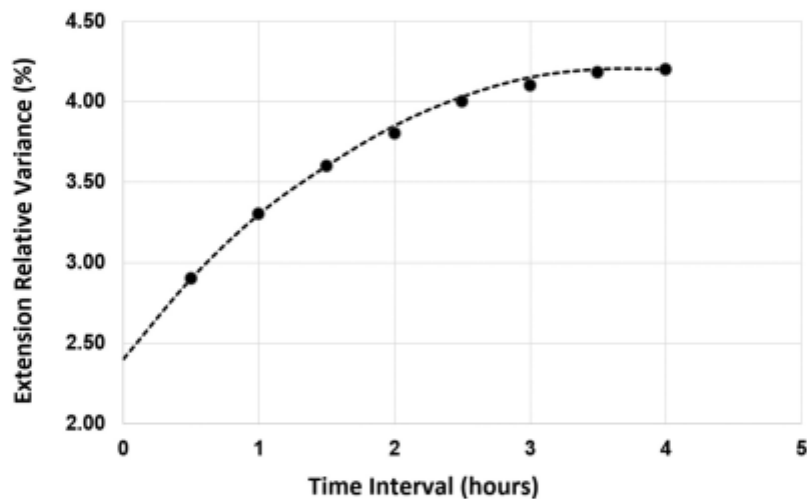


Figure 9. Extension error for ash content.

that occurs when sampling the coal manually, as well as the problem of not knowing the inaccuracy of its sampling process. For both the mining company and the client, it is very important to know the intrinsic variance of the material/lot, as well as the sources of errors associated with its quality assurance and control program.

If possible, this test should even be repeated, to represent all deposits. To reduce costs and time, we can implement geostatistical simulation.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Table 2. Summary of sampling errors.

Sulfur (Average = 0.85%)					Error in % (IC = 95%)	Range (%)	
Time interval	Ext. Var.	Var seq	Lab. Dup. Var.	Total Var.	2 Rel. Stand. Desv.	Min. Lim.	Max. Lim.
30/30 min	0.157	0.008556	0.000625	0.1660	81.48	0.157	1.543
2/2 hours	0.160			0.1687	82.14	0.152	1.548
4/4 hours	0.168			0.1769	84.11	0.135	1.565
Ash (Average = 44%)					Error in % (IC = 95%)	Range (%)	
Time interval	Ext. Var.	Var seq	Lab. Dup. Var.	Total Var.	2 Rel. Stand. Desv.	Min. Lim.	Max. Lim.
30/30 min	0.029000	0.0009	0.000156	0.0301	34.67	28.744	59.256
2/2 hours	0.038000			0.0391	39.53	26.609	61.391
4/4 hours	0.042000			0.0431	41.50	25.740	62.260

ORCID

Gustavo Leandro Tomazi  <http://orcid.org/0000-0002-1196-576X>

References

- Abzalov, M. Z. 2011. Geostatistical approach to the estimation of sampling precision. In: World Conference on Sampling And Blending (WCSB5), 5., 2011, Santiago. Santiago: GECAMIN, 2011. 1, 243–250.
- Abzalov, M. Z. 2016. Applied mining geology. [S.l.]: Springer. *Modern Approaches in Solid Earth Sciences* 12: 244–250.
- Armstrong, M. 1950. *Basic linear geostatistics*. Berlin: Springer-Verlag Berlin Hiedelberg.
- Castilho, M. V., and D. François-bongarçon. 2005. Calibration of parameters for estimating sampling variance. In: World Conference on Sampling and Blending (WCSB2), 2., Melbourne. Melbourne: AusIMM, n.4, 3–8.
- Chierigati, A. C., and F. Pitard. 2009. The challenge of sampling gold. In: World Conference on Sampling and Blending (WCSB4), 4., Cape Town. Cape Town: SAIMM, 2009. 1, 107–12.
- Chierigati, A. C., H. Delboni Jr., J. F. C. L. Costa, and F. B. Carneiro. 2008. Reconciliação pró-ativa em empreendimentos mineiros. *Revista Escola de Minas, [S.l.]*. 61(3):297–302. doi:10.1590/S0370-44672008000300006.
- Deutsch, C. V., and A. G. Journel. 1998. *GSLIB. Geostatistical software library and user's guide*. New York: Oxford University Press.
- Esbensen, K. H., and C. Wagner. 2017a. The variographic experiment: Sampling column. *Spectroscopy Europe, Weinheim* 29: 14–18.
- Esbensen, K. H., and C. Wagner. 2017b. Process sampling: The importance of correct increment extraction. *Spectroscopy Europe, Weinheim* 29 (3):17–20.
- Esbensen, K. H., and K. Engström. 2017. Industrial variographic analysis for continuous sampling system validation. *Spectroscopy Europe, Weinheim* 29 (6):14–17.
- François-Bongarçon, D. 1998. Error variance information from paired data: Applications to sampling theory. *Exploration and Mining Geology, Montreal* 7 (1–2):161–65.
- Gregorieff, A. 2002. Desenvolvimento de um novo protocolo de amostragem de carvão mineral. Masters dissertation, PPGE3M, UFRGS, 2002.
- Gy, P. M. 1979. *Sampling of particulate materials – Theory and practice*. Amsterdam: Elsevier.
- Gy, P. M. 1982. *Sampling of particulate materials – Theory and practice*. Amsterdam: Elsevier.
- Gy, P. M. 1992. *Sampling of heterogeneous and dynamic material systems. Theory of heterogeneity, sampling and homogenizing*. Amsterdam: Elsevier.
- Journel, A. G., and C. J. Huijbregts. 1978. *Mining geostatistics*. New York: Academic Press.
- Marques, D. M., and J. F. C. L. Costa. 2013. An algorithm to simulate ore grade variability in blending and homogenization piles. *International Journal of Mineral Processing, Amsterdam* 120:48–55. doi:10.1016/j.minpro.2013.01.003.