

Comparativo entre o gráfico de estabilidade quantitativo e a regressão linear múltipla para estimativa de diluição em lavras subterrâneas

João Vítor de Souza Nascimento
Universidade Federal de Minas Gerais, Belo Horizonte, Brasil, jvsnascimento@outlook.com

Savio de Oliveira Viana Ribeiro
Engenheiro Geotécnico, AngloGold Ashanti, Sabará, Brasil, saviooliveira96@hotmail.com

RESUMO: O trabalho propõe uma análise comparativa entre o gráfico de estabilidade quantitativo e a regressão linear múltipla para estimativa de diluição em lavras estreitas utilizando o método *sublevel stoping*. O estudo se baseia na relação entre a qualidade do maciço rochoso e a geometria da abertura, conforme descrito por Mathews (1981). A calibração dos intervalos de *overbreak* no gráfico de estabilidade foi realizada por meio de regressão logística, considerando o *Equivalent Radius Factor* (ERF) e o Número de Estabilidade Modificado (N') como variáveis independentes. Os resultados indicam que a escolha do método estatístico deve ser adequada ao comportamento dos dados, visando obter resultados e interpretações precisas e confiáveis no contexto geotécnico. Conclui-se que tanto o gráfico de estabilidade quantitativo quanto a regressão linear múltipla podem ser utilizados em conjunto para estimar a estabilidade dos vãos livres, proporcionando uma abordagem abrangente e confiável.

PALAVRAS-CHAVE: Estimativa de diluição, regressão linear múltipla, regressão logística, geomecânica.

ABSTRACT: The paper proposes a comparative analysis between the quantitative stability chart and multiple linear regression for estimating dilution in narrow mines using the sublevel stoping method. The study is based on the relationship between the quality of the rock mass and the geometry of the opening, as described by Mathews (1981). The calibration of overbreak intervals on the stability graph was carried out using logistic regression, considering the Equivalent Radius Factor (ERF) and the Modified Stability Number (N') as independent variables. The results indicate that the choice of statistical method should be appropriate to the behavior of the data, aiming to obtain precise and reliable results and interpretations in the geotechnical context. It is concluded that both the quantitative stability chart and multiple linear regression can be used together to estimate the stability of voids, providing a comprehensive and reliable approach.

KEYWORDS: Dilution estimation, multiple linear regression, logistic regression, geomechanics.

1 INTRODUÇÃO

Dentre os diversos desafios observados em operações subterrâneas, a estabilidade das escavações é um fator de grande importância, tanto na segurança quanto no controle da diluição operacional, o que motivou o aprofundamento neste estudo. Ao longo dos anos, muitas foram as abordagens desenvolvidas e aprimoradas com o intuito de avaliar a estabilidade dos vãos livres.

Pakalnis (1986) abordou a estimativa de diluição a partir de um estudo de estabilidade de realces. Neste estudo, o objetivo principal consistiu na determinação de diretrizes de estabilidade para a Mina Ruttan. Os resultados mostraram as três principais variáveis de projeto que impactaram na diluição operacional, sendo elas: qualidade do maciço rochoso, taxa de extração e área de superfície exposta no *hangingwall*. Por meio de análises multivariadas, foram elaboradas equações empíricas para estimar a porcentagem de diluição em diferentes grupos de stopes.

Semelhante à abordagem da diluição, Mathews et. al (1981) propuseram o gráfico de estabilidade convencional, que é uma ferramenta qualitativa de design, utilizada para determinar o potencial de estabilidade dos stopes. Para tal, com base em uma relação intuitiva entre a qualidade do maciço rochoso e a geometria da abertura, o gráfico de estabilidade foi proposto, levando em consideração o Raio Hidráulico (RH) e o Número de Estabilidade Modificado (N^*). No estudo em questão, foram considerados 50 casos históricos, que permitiram a definição de zonas de estabilidade, denominadas: Stable Zone, Unstable Zone e Cave Zone.

Mawdesley (2001), utilizou o método da regressão logística para delinear e otimizar o posicionamento das zonas de estabilidade, a partir de um denso banco de dados que compunham o gráfico de estabilidade estendido, contendo mais de 400 casos históricos. Historicamente, as zonas do gráfico tradicional foram definidas visualmente, o que incorpora subjetividade ao método. Sendo assim, a aplicação de um modelo probabilístico é de grande importância, afinal maximiza o valor da metodologia enquanto ferramenta de design, trazendo flexibilidade à otimização das curvas de estabilidade e avaliação da acurácia do modelo. Além disso, conhecendo a probabilidade de estabilidade de um design de escavação, o engenheiro é capaz de considerar o risco e o custo aceitável associado a sobrequebras críticas.

2 CONTEXTO

2.1 Regressão Logística

A regressão logística é uma ferramenta de análise de dados voltada a interpretação de variáveis de natureza dicotômica. Os modelos de regressão logística estimam a probabilidade de ocorrência de um evento, em função das variáveis independentes associadas e das respostas observadas. Neste caso, as variáveis explicativas (independentes) podem ser categóricas ou não e a variável resposta (dependente) é binária.

A probabilidade de ocorrência de um evento pode ser obtida de forma direta. Logo, considerando que a variável dependente Y pode assumir somente dois valores possíveis (1 ou 0) e que o seu resultado pode depender de n variáveis independentes x_1, x_2, \dots, x_n , o modelo de regressão logística é descrito da seguinte forma:

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}} \quad (1)$$

$$g(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2)$$

O termo $P(Y = 1)$ representa a probabilidade de que o evento ocorra. O termo $g(x)$ é denominado *logit*, sendo este uma função das constantes a serem calibradas e das variáveis explicativas do modelo.

O *logit* - o logaritmo natural de uma razão de chances - é o conceito matemático central implícito à regressão logística. Seu resultado representa o logaritmo natural do quociente da probabilidade de ocorrência e não ocorrência do evento, representado pela seguinte equação:

$$\log \frac{P(Y_i=1|X=x)}{P(Y_i=0|X=x)} = \alpha + \beta x_i, \quad i = 1, \dots, n \quad (3)$$

Um valor positivo de *logit* indica que, ao aumentar o valor da variável independente, maior é a probabilidade do evento ocorrer, enquanto um valor negativo de *logit* indica que, o aumento da variável independente está associado a uma menor probabilidade do evento ocorrer.

O modelo é calibrado a partir do método da Máxima Verossimilhança, que retorna os valores ótimos das constantes α e β . Os estimadores de máxima verossimilhança de α e β correspondem a solução das seguintes equações (Morentin & Singer, 2021):

$$\sum_{i=1}^n \left\{ y_i - \frac{\exp(\hat{\alpha} + \hat{\beta}x_i)}{1 + \exp(\hat{\alpha} + \hat{\beta}x_i)} \right\} = 0 \quad \sum_{i=1}^n x_i \left\{ y_i - \frac{\exp(\hat{\alpha} + \hat{\beta}x_i)}{1 + \exp(\hat{\alpha} + \hat{\beta}x_i)} \right\} = 0 \quad (4)$$

2.2 Regressão Linear Múltipla

O modelo de regressão linear múltipla é descrito em termos de n variáveis explicativas x_1, \dots, x_n , que podem ser contínuas ou categóricas, e uma variável resposta contínua, como mostrado a seguir:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i, \quad i = 1, \dots, n \quad (5)$$

Neste caso, testar hipóteses dos coeficientes do modelo ou construir intervalos de confiança para eles, exige supor que a distribuição de frequências dos erros é Normal. Sendo assim, o modelo possui $p+2$ parâmetros desconhecidos, nomeadamente, $\beta_0, \beta_1, \dots, \beta_p$ e σ^2 , que devem ser estimados com base nos dados observados.

O modelo de regressão linear múltipla é calibrado a partir da minimização do somatório dos erros ao quadrado, como mostra a equação a seguir:

$$Q(\beta_0, \dots, \beta_p) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - \sum_{j=0}^p \beta_j x_{ij}]^2 \quad (6)$$

Em linhas gerais, a operação consiste em encontrar os valores ótimos de $\beta_0, \beta_1, \dots, \beta_p$, que minimizem o resultado do somatório dos erros ao quadrado. O erro é a diferença entre o valor real observado e o valor estimado para diversas hipóteses de coeficientes, como indicado a seguir:

$$\hat{e} = y_i - \hat{y}_i, \quad i = 1, \dots, n \quad (7)$$

3 METODOLOGIA

3.1 Coleta de Dados

A coleta de dados para calibração do gráfico de estabilidade por regressão logística, compreendeu dois diferentes corpos da mina. No total foram utilizados 30 dados, sendo 14 do Corpo Q e 16 do Corpo C. Optou-se por substituir o Raio Hidráulico (RH) pelo *Equivalent Radius Factor* (ERF), por ser mais flexível em stopes com geometrias mais complexas e por ser facilmente calculado utilizando-se o software Gem4D.

$$ERF = \frac{0,5}{1/n \sum_{\theta=1}^n 1/r_{\theta}} \quad (8)$$

Como mencionado, utilizou-se o software Gem4D para o cálculo do ERF na face crítica de cada vão formado, como exemplifica a imagem abaixo:

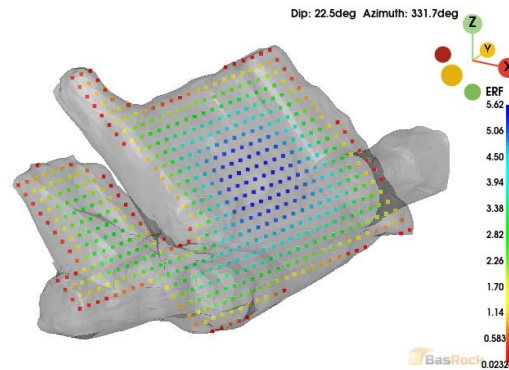


Figura 1. ERF na face *hangingwall* em stope de lavra.

Sequencialmente, realizou-se o levantamento do volume de *overbreak* no *hangingwall* dos stopes por meio do software Datamine RM, bem como do Número de Estabilidade Modificado (N’).

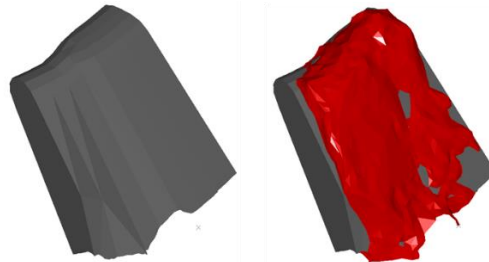


Figura 2. *Overbreak* na face *hangingwall* em stope de lavra.

Os valores de N’ foram obtidos a partir de mapeamentos realizados em campo, e refletem as particularidades de cada região de lavra, tais quais atitudes das principais famílias de discontinuidades, contatos litológicos, entre outros parâmetros. Para mais informações acerca do N’ e seus fatores, consultar Mathews (1981). Segue abaixo um exemplo dos mapeamentos geológicos utilizados:

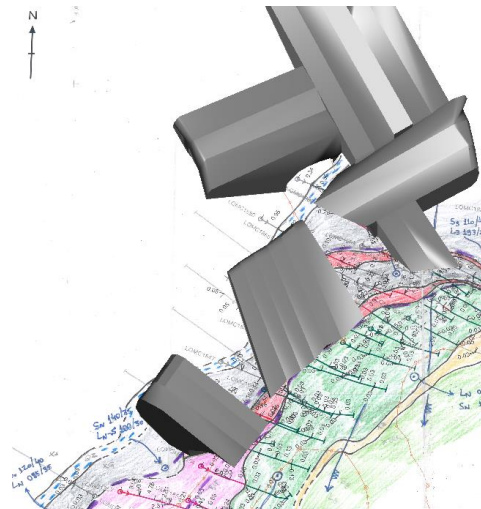


Figura 3. Mapeamento de galerias

De posse do banco de dados, os stopes foram classificados de acordo seu percentual de *overbreak*, se enquadrando nas seguintes classes:

- CLASSE A: *Overbreak* < 7%
- CLASSE B: *Overbreak* 7 - 14%
- CLASSE C: *Overbreak* > 14%

3.2 Aplicação da Regressão Logística

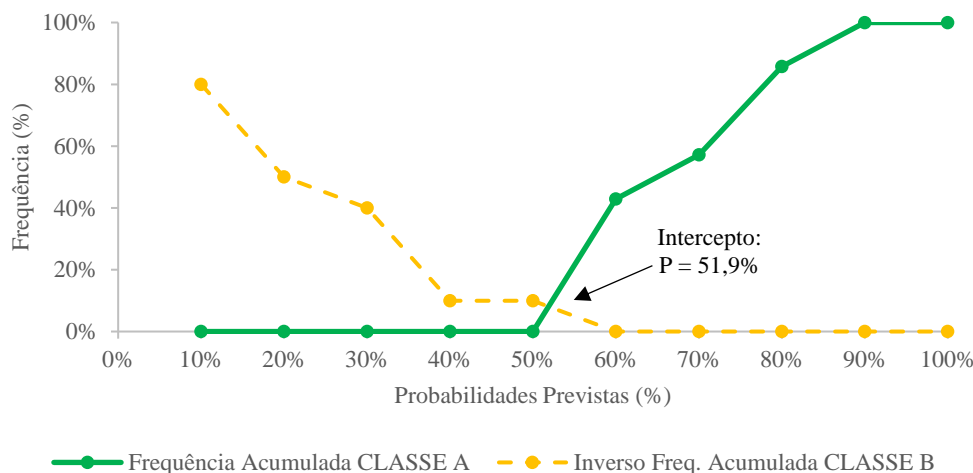
A regressão logística foi realizada utilizando-se linguagem de programação Python no ambiente Google Colab. A definição das zonas/limites de estabilidade foram feitas separadamente, primeiro entre a Classe A e a Classe B, em seguida entre as Classes A e B, e os dados da Classe C. Logo, para cada um dos conjuntos de dados foi obtida uma função *logit*, contendo os valores calibrados dos coeficientes de regressão, α , β_1 e β_2 (Equação 9).

$$z = \alpha + \beta_1 \ln ERF + \beta_2 \ln N' \quad (9)$$

Experimentos com este tipo de banco de dados mostraram que a transformação logarítmica dos termos N' e RH (Neste caso ERF) fornece uma resposta melhor, gerando limites de separação lineares para as classes de estabilidade (Mawdesley, 2002). Logo, os dados de entrada da regressão logística foram dados em termos do logaritmo natural.

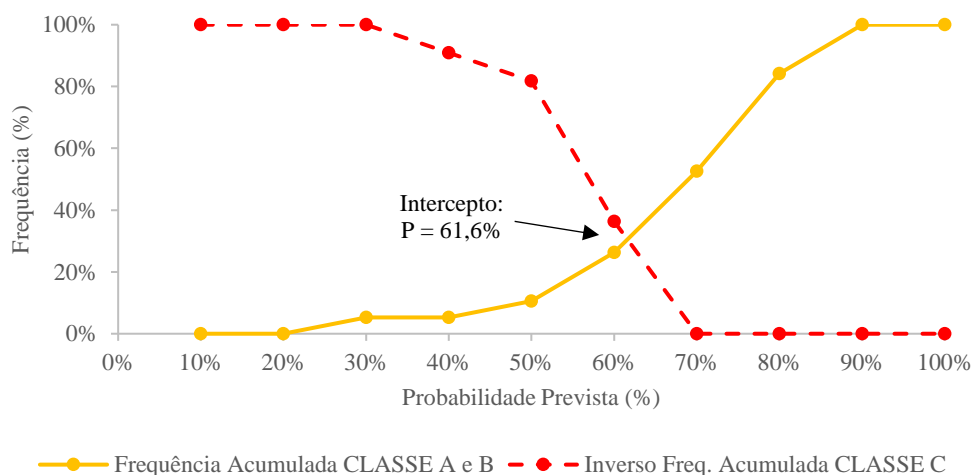
O próximo passo consistiu na delimitação da melhor posição de cada um dos limites de classes. Para tal, é necessário levar em consideração a frequência acumulada das probabilidades previstas pelo modelo *logit*, bem como o inverso da frequência acumulada das probabilidades da classe adjacente. Sendo assim, a ideia consiste em obter o valor da probabilidade prevista na interseção entre a curva de frequência acumulada da Classe A e a curva do inverso da frequência acumulada da Classe B, e assim sucessivamente.

Gráfico 1. Frequência acumulada das probabilidades previstas entre a Classe A e B



O mesmo foi realizado entre os dados das Classes A, B e a Classe C.

Gráfico 2. Frequência acumulada das probabilidades previstas entre a Classe AB e C



Existem vantagens em delinear matematicamente os limites entre classes, quando comparado ao método meramente visual, principalmente por trazer mais objetividade e a possibilidade de quantificação da variância dos limites estimados. Deste modo, os limites entre classes são expressos matematicamente da seguinte forma (Mortazavi, 2021).

$$N' = e^{\frac{(z - \alpha - \beta_1 \ln ERF)}{\beta_2}} \quad (10)$$

Segue o desenvolvimento matemático da igualdade:

$$P(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta_1 \ln ERF + \beta_2 \ln N')}}}$$

$$\frac{1}{P} = 1 + e^{-(\alpha + \beta_1 \ln ERF + \beta_2 \ln N')}$$

$$\frac{1}{P} - 1 = e^{-(\alpha + \beta_1 \ln ERF + \beta_2 \ln N')}$$

$$\ln\left(\frac{1}{P} - 1\right) = -(\alpha + \beta_1 \ln ERF + \beta_2 \ln N')$$

$$\ln\left(\frac{1}{P} - 1\right) + \alpha + \beta_1 \ln ERF = -\beta_2 \ln N'$$

$$\beta_2 \ln N' = -\ln\left(\frac{1}{P} - 1\right) - \alpha - \beta_1 \ln ERF$$

$$\ln N' = \frac{-\ln\left(\frac{1}{P} - 1\right) - \alpha - \beta_1 \ln ERF}{\beta_2}$$

$$N' = e^{\frac{-\ln\left(\frac{1}{P} - 1\right) - \alpha - \beta_1 \ln ERF}{\beta_2}}$$

$$N' = e^{\frac{(z - \alpha - \beta_1 \ln ERF)}{\beta_2}} \quad (11)$$

O parâmetro z , mostrado na Equação 11, é o valor de *logit* resultante da probabilidade na interseção dos gráficos 1 e 2. Logo, cada limite foi construído seguindo a Equação 11 e considerando a probabilidade na interseção entre as respectivas classes. A estimativa do valor *logit* para a definição dos limites é calculada da seguinte forma:

$$z = \ln\left(\frac{P}{1-P}\right) = -\ln\left(\frac{1}{P} - 1\right) \quad (12)$$

No qual,

z : Valor do *logit* previsto

p : Valor da probabilidade prevista pelo modelo *logit* na interseção

3.3 Aplicação da Regressão Linear Múltipla

Utilizou-se o software Excel e a ferramenta Análise de Dados para realização da regressão. Para a análise de regressão linear múltipla foram considerados os mesmos 30 dados utilizados na regressão logística, sendo a porcentagem de *overbreak* a variável dependente e N' e ERF as variáveis independentes. Considerando-se os resultados obtidos na tabela abaixo, referentes à calibração do modelo de regressão linear múltipla, observou-se que os resultados do teste de significância das variáveis (valor-P) podem não ser confiáveis, visto que a relação linear não é tão evidente entre os valores de porcentagem de *overbreak* e ERF, ou seja, o valor p-value pode ser questionado e a hipótese nula menos provável.

Tabela 1. Resultados da regressão linear múltipla

	Coefficientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores	Inferior 95.0%	Superior 95.0%
Interseção	0.1362	0.0655	2.0807	0.0471	0.0019	0.2705	0.0019	0.2705
N'	-0.0036	0.0013	-2.7896	0.0096	-0.0062	-0.0009	-0.0062	-0.0009
ERF	0.0070	0.0114	0.6156	0.5433	-0.0164	0.0305	-0.0164	0.0305

A regressão linear múltipla abre margem à observação de diversas outras variáveis geotécnicas, que podem ser utilizadas a fim de estimar a sobrequebra em faces de stopes. Para mais informações acerca deste método, consultar Costa (2017).

4 RESULTADOS E DISCUSSÕES

4.1 Resultados da Regressão Logística

Limite Superior:

Acurácia do modelo: 0.894 (17 acertos de 19 dados)

Coefficientes da regressão: [[-0.88249725 1.58715811]]

Coefficiente de interceptação: [-3.4661699]

$$N' = e^{\frac{[z - (-3,466) - (-0,882)\ln\text{ERF}]}{1,587}} \quad (13)$$

$$z = \ln\left(\frac{0,519}{1-0,519}\right) \quad (14)$$

Limite Inferior:

Acurácia do modelo: 0.733 (22 acertos de 30 dados)

Coefficientes da regressão: [[-0.45958349 0.74319632]]

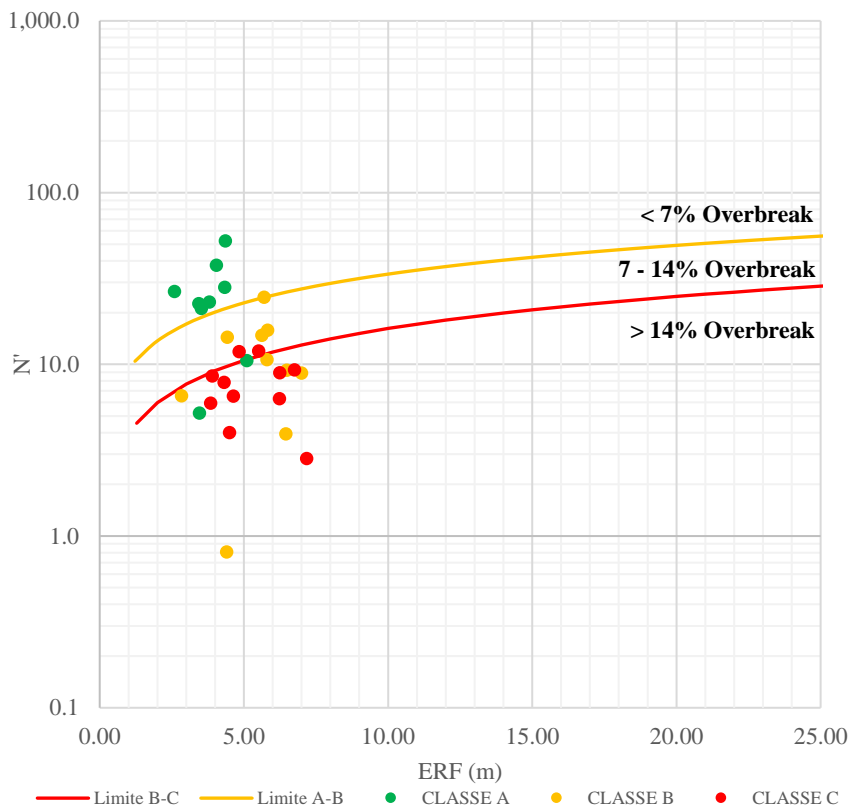
Coefficiente de interceptação: [-0.42941662]

$$N' = e^{\frac{[z - (-0,429) - (-0,459)\ln\text{ERF}]}{0,743}} \quad (15)$$

$$z = \ln\left(\frac{0,616}{1-0,616}\right) \quad (16)$$

De acordo o valor de acurácia do modelo do limite superior, 89,4% das observações foram classificadas corretamente nas Classes A e B, enquanto no limite inferior, 73,3% das observações foram classificadas corretamente nas Classes AB e C, ou seja, representam a taxa de acerto no modelo dentro dos limites de *overbreak* propostos. É válido salientar que, a acurácia pode não ser suficiente para determinar a qualidade do modelo, principalmente em conjuntos de dados onde uma classe é muito mais frequente do que outra. Logo, os limites foram traçados a partir da equações 13 e 15, como mostra o gráfico a seguir:

Gráfico 3. Gráfico de estabilidade quantitativo calibrado



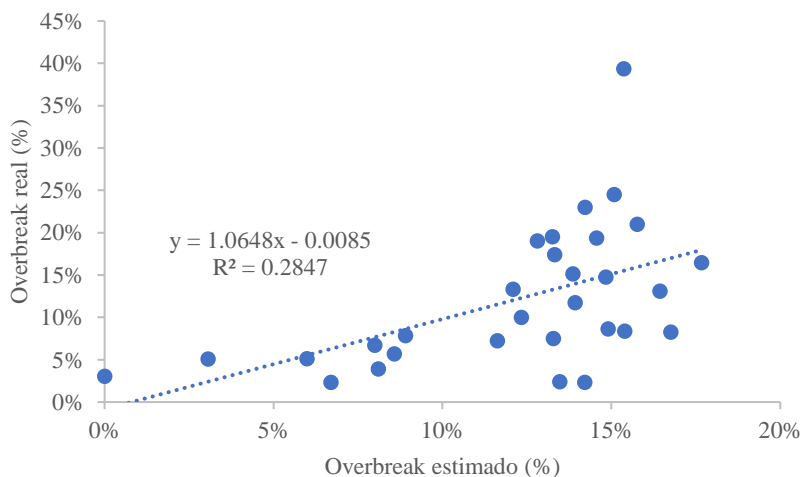
4.2 Resultados da Regressão Linear Múltipla

A partir dos valores de coeficientes retornados (Tabela 1) obteve-se a seguinte equação preditiva de porcentagem de *overbreak*:

$$OB (\%) = (0,1362 + 0,007ERF - 0,0036N') \cdot 100 \quad (17)$$

Logo, a porcentagem de *overbreak* foi estimada para a base de dados usada na calibração. Os resultados mostraram $R^2 = 0.2847$

Gráfico 4. Dispersão entre valores de *overbreak* estimado e real



5 CONSIDERAÇÕES FINAIS

De acordo os resultados observados, a escolha do método estatístico deve estar alinhada ao comportamento dos dados, a fim de retornar resultados e interpretações precisas e confiáveis no contexto geotécnico. Neste caso, a regressão logística foi considerada mais confiável, pois classificou com acurácia os realces dentro dos intervalos de *overbreak* propostos. No entanto, como mencionado, a acurácia do modelo pode não ser suficiente para determinar sua qualidade, principalmente em conjuntos de dados desbalanceados, ou seja, que contém classes com frequências de observações maiores do que outras. Neste caso, o balanceamento do modelo foi realizado a partir dos gráficos 1 e 2, o que melhorou sua taxa de acertos. Na regressão logística existem outras formas de avaliar a qualidade do modelo, além da acurácia, sendo elas o F1-Score e o Recall. A acurácia indica a proporção de previsões corretas, tanto acima quanto abaixo dos limites do gráfico de estabilidade, ou seja, previsões corretas de verdadeiros positivos e verdadeiros negativos. É considerada uma métrica fácil, intuitiva e confiável em classes balanceadas.

Além disso, a partir da regressão logística é possível delinear no gráfico de estabilidade curvas de iso-probabilidade, que ilustram com facilidade a probabilidade de ocorrência para cada classe utilizada, em função dos valores de N' e ERF. A utilização de curvas de iso-probabilidade é uma boa alternativa para classificação de risco de lavra, possibilitando julgar quais classes de *overbreak* são mais prováveis de acontecer em um determinado realce.

Por outro lado, a regressão linear múltipla também é capaz de descrever o comportamento de sobrequebras em realces, o que depende das variáveis geotécnicas selecionadas e, principalmente, de um banco de dados sólido e confiável. No entanto, o método carrega consigo algumas limitações, principalmente por assumir uma relação de linearidade entre as variáveis independentes e a variável dependente. Considerando uma relação não linear, o modelo pode resultar em previsões imprecisas que não descrevem a dinâmica dos dados. Outrossim, a regressão linear múltipla é mais sensível a *outliers*, o que prejudica a qualidade do modelo.

A utilização de cada um dos métodos depende do tipo de resultado esperado para a variável dependente. Para previsões contínuas de porcentagem de *overbreak* a regressão linear múltipla pode ser mais adequada. Para classificação binária e avaliação probabilística de risco, a regressão logística é a escolha apropriada. Portanto, ambos os métodos podem ser utilizados em consonância para estimar a estabilidade dos vãos livres de forma eficaz, o que depende do comportamento dos dados geotécnicos, das variáveis explicativas selecionadas e da qualidade do banco de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Costa, L. C. B. (2017). *Metodologia para prever Costa, e reduzir o risco de diluição não planejada em lavras estreitas pelo método Sublevel Stopping. Estudo de caso: Mina Córrego do Sítio, Santa Bárbara/MG.* Universidade Federal de Ouro Preto, Minas Gerais, 106 p.
- Mathews, K.; Hoek, E.; Wyllie, D. C.; Stewart, S. B. V. (1981). *Prediction Of Stable Excavation Spans for Mining at Depths below 1000 Meters in Hard Rock*, Report to Canada Centre for Mining and Energy Technology, 110 p.
- Mawdesley, C. A. (2002). *Predicting rock mass cavability in block caving mines.* PhD thesis (unpublished), Julius Kruttschnitt Mineral Research Centre. University of Queensland, Brisbane.
- Mawdesley, C. A.; Trueman, R.; Whiten, W. (2001). *Extending the Mathews stability graph for Open-stope design, Transactions of the Institutions of Mining and Metallurgy, Mining Technology*, 110(1):A27–A39.
- Morettin, P. A.; Singer, J. M. (2021). *Estatística e Ciência de Dados.* LTC, São Paulo, 489 p.
- Mortazavi, A.; Osserbay, B. (2021). *The Consolidated Mathews Stability Graph for Open Stope Design.* Kazakhstan, Nazarbayev University, 22 p.
- Pakalnis, R. T. (1986). *Empirical stope design at the Ruttan Mine, Sherritt Gordon Mines Ltd.* University of British Columbia, 276 p. DOI: <https://dx.doi.org/10.14288/1.0081095>