



UNIVERSIDADE FEDERAL DE OURO PRETO
ESCOLA DE MINAS
DEPARTAMENTO DE ENGENHARIA DE MINAS



José Augusto Oliveira da Rocha

**Aplicação de Regressão Linear Múltipla na Estimativa do Fator de Segurança
de Taludes Sujeitos à Ruptura Circular**

**Ouro Preto
2026**

José Augusto Oliveira da Rocha

**Aplicação de Regressão Linear Múltipla na Estimativa do Fator de Segurança
de Taludes Sujeitos à Ruptura Circular**

Trabalho de Conclusão de Curso apresentado à
Universidade Federal de Ouro Preto como parte
dos requisitos para obtenção do Grau de
Engenharia de Minas.

Orientadora: Prof.^a Dra. Tatiana Barreto dos
Santos

Ouro Preto

2026



FOLHA DE APROVAÇÃO

José Augusto Oliveira da Rocha

Aplicação de Regressão Linear Múltipla na Estimativa do Fator de Segurança de Taludes Sujeitos à Ruptura Circular

Monografia apresentada ao Curso de Engenharia de Minas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Engenharia de Minas

Aprovada em 06 de fevereiro de 2026

Membros da banca

Drª Tatiana Barreto dos Santos - Orientador(a) - Universidade Federal de Ouro Preto
Dr. Felipe Ribeiro Souza - Universidade Federal de Ouro Preto
Dr. José Fernando Miranda - Universidade Federal de Ouro Preto

Tatiana Barreto dos Santos, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 10/02/2026



Documento assinado eletronicamente por **Tatiana Barreto dos Santos, PROFESSOR DE MAGISTERIO SUPERIOR**, em 10/02/2026, às 08:24, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1056172** e o código CRC **C034A1CA**.

Agradecimentos

Agradeço primeiramente a Deus, pois foi Ele quem me deu forças para chegar até aqui.

Agradeço aos meus pais, Alan e Regina, por todo o amor, apoio e por me inspirarem diariamente. Agradeço também aos meus avós, Carlos (em memória) e Geralda (em memória), por serem meus exemplos na vida.

Agradeço à lara por todo o amor, carinho e apoio durante esta etapa.

Agradeço aos amigos que estiveram comigo ao longo desta caminhada.

À República Partenon e a todos os seus moradores e ex-alunos. Vocês me proporcionaram um lar em Ouro Preto; serei sempre grato por tudo o que vivi nos últimos anos.

Agradeço à UFOP, à Escola de Minas e ao DEMIN por todo o aprendizado.

Por fim, agradeço à Prof.^a Tatiana pelos conselhos e pelo apoio durante a realização deste trabalho.

Resumo

A estabilidade de taludes é um tema importante da engenharia geotécnica, sendo o Fator de Segurança (FS) o principal indicador do risco de ruptura. Este trabalho avaliou a aplicação de modelos de regressão linear múltipla para estimar o FS de taludes sujeitos à ruptura circular, analisando a influência do tratamento de dados no desempenho e na confiabilidade estatística dos modelos. Foi utilizado um banco de dados contendo parâmetros geotécnicos e geométricos de taludes. Inicialmente, realizou-se análise exploratória para identificação de inconsistências, com remoção de observações fisicamente incoerentes, reduzindo o conjunto de 87 para 71 amostras válidas. Em seguida, foram definidos três cenários de modelagem: (i) modelo com dados brutos; (ii) modelo com remoção de *outliers* multivariados por meio da distância de Mahalanobis; e (iii) modelo com dados tratados segundo critérios de valores típicos das variáveis independentes. A modelagem foi conduzida por regressão linear múltipla, acompanhada de análise gráfica, avaliação de resíduos e validação cruzada K-fold no terceiro cenário. A análise dos resíduos indicou comportamento aleatório, sem tendência ou estrutura, indicando que o ajuste para os três cenários foi satisfatório. Os resultados mostraram que o modelo com dados brutos apresentou baixo poder explicativo, enquanto a remoção de *outliers* no segundo cenário resultou em melhora moderada. O melhor desempenho foi obtido no terceiro cenário com tratamento de dados mais robusto, que apresentou elevado R² ajustado e bom desempenho preditivo na validação cruzada. Conclui-se que o tratamento prévio dos dados influencia de forma decisiva o desempenho de modelos de regressão linear múltipla aplicados à estabilidade de taludes. O modelo baseado em valores típicos mostrou maior confiabilidade estatística, porém com restrição de aplicabilidade, devendo ser utilizado como ferramenta complementar às análises determinísticas tradicionais.

Palavras-chave: Estabilidade de taludes; Fator de Segurança; Regressão linear múltipla; Tratamento de dados; Geotecnia.

Abstract

Slope stability is an important topic in geotechnical engineering, with the Factor of Safety (FS) being the main indicator of failure risk. This study evaluated the application of multiple linear regression models to estimate the FS of slopes subjected to circular failure, analyzing the influence of data treatment on model performance and statistical reliability. A database containing geotechnical and geometric slope parameters was used. Initially, an exploratory analysis was carried out to identify inconsistencies, and physically incoherent observations were removed, reducing the dataset from 87 to 71 valid samples. Subsequently, three modeling scenarios were defined: (i) model using raw data; (ii) model with removal of multivariate outliers using Mahalanobis distance; and (iii) model with data treated according to typical value criteria for the independent variables. Modeling was performed using multiple linear regression, accompanied by graphical analysis, residual evaluation, and K-fold cross-validation in the third scenario. Residual analysis indicated random behavior, with no trend or structure, suggesting satisfactory fit across the three scenarios. The results showed that the model with raw data had low explanatory power, while outlier removal in the second scenario led to moderate improvement. The best performance was obtained in the third scenario, with more robust data treatment, which presented a high adjusted R^2 and good predictive performance in cross-validation. It is concluded that prior data treatment decisively influences the performance of multiple linear regression models applied to slope stability. The model based on typical values showed greater statistical reliability; however, its applicability is limited and it should be used as a complementary tool to traditional deterministic analyses.

Keywords: Slope stability; Factor of Safety; Multiple linear regression; Data treatment; Geotechnics.

Lista de Figuras

Figura 1 - Terminologia aplicada a geometria de taludes.	13
Figura 2 - Critério de Ruptura de Mohr-Coulomb.	14
Figura 3 - Valores típicos de coesão.	15
Figura 4 - Valores típicos de ângulo de atrito para diferentes tipos de rocha.	16
Figura 5 - Peso Específico de Solos Argilosos.	17
Figura 6 - Peso Específico de Solos Arenosos.	17
Figura 7 - Relação do Fator de Segurança.	18
Figura 8 - Ruptura Circular.	19
Figura 9 - Boxplot Geral – Análise 1.	30
Figura 10 - Histogramas - Análise 1.	30
Figura 11 - Matriz de dispersão – Análise 1.	31
Figura 12 - Dispersão dos resíduos – Análise 1.	32
Figura 13 - Histograma dos resíduos – Análise 1.	33
Figura 14 - Distâncias de Mahalanobis e limite crítico – Análise 2.	35
Figura 15 - Dispersão dos Resíduos – Análise 2.	37
Figura 16 - Histograma dos resíduos – Análise 2.	37
Figura 17 - Boxplot Geral – Análise 3.	40
Figura 18 - Histogramas - Análise 3.	41
Figura 19 - Matriz de dispersão – Análise 3.	42
Figura 20 - Distâncias de Mahalanobis e limite crítico - Análise 3.	43
Figura 21 - Dispersão dos Resíduos – Análise 3.	44
Figura 22 - Histograma dos Resíduos – Análise 3.	44

Lista de Tabelas

Tabela 1 - Resultados Teste de Bartlett	28
Tabela 2 - Resumo Estatística Descritiva – Análise 1.....	29
Tabela 3 - Resíduos – Análise 1.....	32
Tabela 4 - Coeficientes de Regressão – Análise 1.....	33
Tabela 5 - Resumo Estatística Descritiva pós remoção de outliers – Análise 2.....	36
Tabela 6 - Resíduos – Análise 2.....	36
Tabela 7 - Coeficiente de Regressão – Análise 2.....	38
Tabela 8 - Resumo Estatística Descritiva - Análise 3.....	40
Tabela 9 - Resíduos – Análise 3.....	43
Tabela 10 - Coeficiente de Regressão – Análise 3.....	45
Tabela 11 - Resultados da Validação Cruzada.....	46
Tabela 12 - Comparação de resultados dos modelos.....	48

Sumário

1. INTRODUÇÃO	10
2. OBJETIVOS	11
2.1 Objetivo Geral	11
2.2 Objetivos Específicos	11
3. JUSTIFICATIVA	11
4. FUNDAMENTAÇÃO TEÓRICA	12
4.1 Estabilidade de Taludes	12
4.2 Resistência ao Cisalhamento	13
4.2.1 Coesão	14
4.2.2 Ângulo de Atrito	15
4.2.3 Razão de Poropressão	16
4.3 Peso Específico	17
4.4 Método do Equilíbrio Limite	18
4.4.1 Fator de Segurança	18
4.5 Ruptura Circular.....	19
4.6 Estatística Multivariada.....	20
4.6.1 Correlação entre variáveis	21
4.6.1.1 Fórmula do coeficiente de correlação de Pearson.....	21
4.6.2 Regressão linear múltipla.....	21
4.6.2.1 Forma geral do modelo	21
4.6.3 Teste de esfericidade de Bartlett.....	22
4.6.4 Detecção de <i>outliers</i> multivariados	23
5. MATERIAIS E MÉTODOS	23
5.1 Banco de Dados	23
5.1.1 Variáveis do estudo	23
5.2 Ambiente computacional	24
5.2.1 RStudio	24
5.3 Análise exploratória dos dados	24
5.4 Testes estatísticos preliminares	25
5.4.1 Teste de Esfericidade de Bartlett	25
5.5 Identificação e tratamento de <i>outliers</i>.....	25

5.6 Construção dos modelos de regressão	26
5.6.1 Critérios adotados para tratamento dos dados na Análise 3	26
5.7 Análise dos modelos.....	27
5.8 Validação dos modelos.....	27
6. RESULTADOS E DISCUSSÕES	27
6.1 Análise Preliminar	28
6.2 Cenários de Análise	28
6.3 Teste de Bartlett	28
6.4 Análise 1	29
6.4.1 Estatística descritiva	29
6.4.2 Correlações entre as variáveis.....	31
6.4.3 Desempenho do modelo.....	32
6.4.3.1 Coeficientes de Regressão	33
6.4.3.2 Fórmula do modelo de regressão.....	34
6.4.4 Avaliação do modelo	34
6.5 Análise 2	35
6.5.1 Identificação de <i>outliers</i>	35
6.5.2 Estatística Descritiva pós remoção de <i>outliers</i>	36
6.5.3 Desempenho do modelo.....	36
6.5.3.1 Coeficientes de regressão	38
6.5.3.2 Fórmula do modelo de regressão.....	39
6.5.4 Avaliação do modelo	39
6.6 Análise 3	39
6.6.1 Estatística descritiva	39
6.6.2 Correlações entre as variáveis.....	42
6.6.3 Identificação de <i>outliers</i>	42
6.6.4 Desempenho do modelo.....	43
6.6.4.1 Coeficientes de regressão	45
6.6.4.2 Fórmula do modelo de regressão.....	46
6.6.5 Validação Cruzada	46
6.6.6 Avaliação do modelo	47
6.7 Síntese das Análises.....	48

7. CONCLUSÕES	49
REFERÊNCIAS BIBLIOGRÁFICAS.....	50
APÊNDICES	53
ANEXOS.....	67

1. INTRODUÇÃO

A estabilidade de taludes é um dos principais desafios da engenharia geotécnica, estando diretamente relacionada à segurança em ambientes de mineração e obras civis. Rupturas de taludes podem resultar em perdas econômicas significativas, impactos ambientais e consequentemente, riscos à vida humana. Sendo assim, a correta avaliação da estabilidade é etapa fundamental em projetos que envolvem escavações em taludes artificiais ou encostas.

Tradicionalmente, a análise da estabilidade de taludes é realizada por meio de métodos determinísticos. Segundo Gerscovich (2016) esses métodos permitem a obtenção do Fator de Segurança (FS) por meio da comparação entre as tensões cisalhantes mobilizadas e a resistência ao cisalhamento do maciço. Contudo, para sua aplicação, é necessária a definição prévia de alguns parâmetros geotécnicos, tais como peso específico, coesão, ângulo de atrito, geometria do talude e condições de poropressão.

Entretanto, a obtenção de parâmetros confiáveis nem sempre é simples, uma vez que esses dados são oriundos de pesquisas de investigação, ensaios laboratoriais, retroanálises e interpretações empíricas. Nesse contexto, a busca por ferramentas auxiliares a avaliação preliminar da estabilidade torna-se relevante.

Com o avanço das ferramentas computacionais e da análise de dados, métodos estatísticos e técnicas multivariadas têm sido cada vez mais utilizados na engenharia. Dentre essas técnicas, a regressão linear múltipla se destaca como uma abordagem capaz de modelar relações entre uma variável dependente e múltiplas variáveis independentes, permitindo avaliar, simultaneamente, a influência dos parâmetros no fator de segurança (FS) em taludes.

Diante desse cenário, o presente trabalho propõe o desenvolvimento e a avaliação de modelos de regressão linear múltipla para estimativa do Fator de Segurança. O banco de dados utilizado foi compilado e organizado por Ahmad et al (2022), ele contempla 87 taludes reais oriundos da literatura, sujeitos, principalmente, a ruptura circular, caracterizados por sete parâmetros geomecânicos.

São analisados diferentes cenários de tratamento dos dados, incluindo a remoção de *outliers* e a aplicação de critérios baseados em valores típicos dos

parâmetros geomecânicos. O objetivo é avaliar o impacto dessas etapas na qualidade estatística e na confiabilidade dos modelos gerados.

É importante ressaltar que essa análise constitui uma ferramenta complementar aos métodos clássicos de análise de estabilidade de taludes. Conforme mencionado por Santos (2016), os métodos estatísticos multivariados não substituem as análises tradicionais de estabilidade de taludes, entretanto, podem contribuir para a identificação preliminar de taludes que demandam de uma investigação mais detalhada.

2. OBJETIVOS

2.1 Objetivo Geral

Desenvolver e avaliar modelos de regressão linear múltipla para estimativa do Fator de Segurança de taludes sujeitos a ruptura circular.

2.2 Objetivos Específicos

Abaixo estão descritos os objetivos específicos do presente trabalho:

- Analisar estatisticamente o banco de dados;
- Avaliar diferentes cenários de tratamento de dados;
- Verificar a adequação da regressão linear;
- Comparar os modelos desenvolvidos;
- Validar os modelos por validação cruzada;

3. JUSTIFICATIVA

A análise de estabilidade de taludes é fundamental na engenharia geotécnica, não apenas para garantir a segurança de obras e vidas humanas, mas também para mitigar riscos concretos em áreas vulneráveis.

De acordo com o Serviço Geológico do Brasil – SGB (CPRM, 2021), o município de Ouro Preto (MG) apresenta o maior número de áreas de risco geológico no país,

com 313 áreas mapeadas das quais cerca de 97% são classificadas como de risco geológico alto e 2,8% como muito alto, atingindo diretamente mais de 3 mil pessoas.

Ainda segundo o SGB (CPRM, 2021), 882 domicílios particulares e coletivos estão localizados nestas áreas de risco, e mais de 80% desses domicílios se situam em áreas suscetíveis a deslizamentos de encostas, evidenciando a relevância deste problema.

A aplicação de modelos estatísticos na estabilidade de taludes representa uma ferramenta complementar para análises preliminares, retroanálises e estudos comparativos, especialmente quando há dados históricos disponíveis. Modelos de regressão linear múltipla permitem estimar o fator de segurança (FS) em taludes de forma rápida e apoiam na tomada de decisão em diversos projetos de engenharia.

Este trabalho analisa o desempenho de regressões lineares para estimar o FS e demonstra a influência do tratamento dos dados na confiabilidade dos modelos. Ao mostrar as vantagens, as limitações e as condições em que esses modelos podem ser aplicados, o estudo contribui para o uso cauteloso de técnicas de análise de dados na geotecnia, de forma complementar aos métodos determinísticos tradicionais de análise de estabilidade.

4. FUNDAMENTAÇÃO TEÓRICA

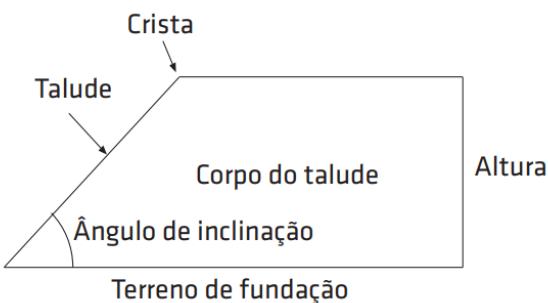
4.1 Estabilidade de Taludes

A avaliação da estabilidade de taludes, fundamenta-se nas tensões cisalhantes atuantes e na sua comparação com a resistência ao cisalhamento do solo, ambas ao longo de uma superfície de ruptura considerada crítica. De acordo com Gerscovich (2016), a ruptura ocorre com a formação de uma superfície de cisalhamento contínua na massa do solo. Nesse processo, uma camada de solo ao redor dessa superfície perde parte de suas características, formando a zona cisalhada.

Gerscovich (2016) e Fiori (2015), definem talude como qualquer superfície inclinada que limita um maciço de solo ou rocha. Os taludes podem ser classificados como naturais, também denominados encostas, ou artificiais, quando resultantes das intervenções do homem.

De acordo com Santos (2016), a altura e o ângulo de inclinação do talude constituem parâmetros fundamentais nos estudos de estabilidade, uma vez que são variáveis que podem ser alteradas quando as medidas adotadas contribuem para a instabilidade do talude. Na figura 1, são representados alguns termos principais para geometria de taludes.

Figura 1 - Terminologia aplicada a geometria de taludes.



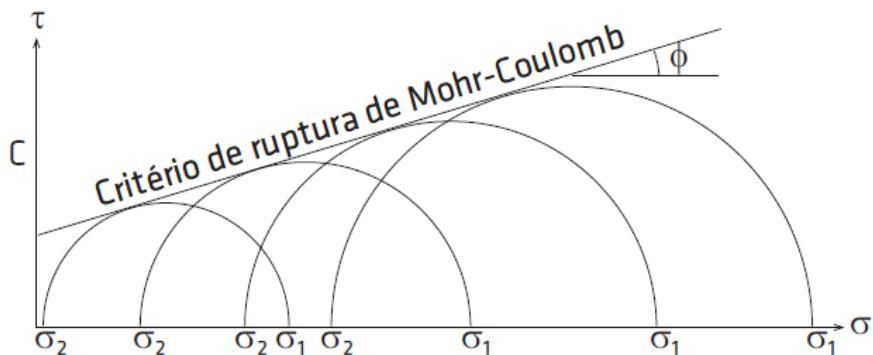
Fonte: Fiori (2015).

4.2 Resistência ao Cisalhamento

A ruptura por cisalhamento caracteriza-se pelo desenvolvimento de uma superfície de ruptura na qual a tensão cisalhante atinge um valor crítico, sendo seguida pelo deslocamento ao longo do plano de ruptura e pelo relaxamento dos esforços atuantes (FIORI, 2015).

Ainda segundo Fiori (2015), o critério de ruptura mais simples e amplamente difundido é o critério de Mohr–Coulomb, o qual é representado por uma envoltória linear tangente ao círculo de Mohr, definindo as condições críticas associadas às combinações dos esforços principais. Na figura 2, é exibida a representação gráfica do critério de ruptura de Mohr-Coulomb.

Figura 2 - Critério de Ruptura de Mohr-Coulomb.



Fonte: Fiori (2015).

A equação da reta apresentada na figura 2 é dada por:

$$\tau = c + \sigma \operatorname{tg} \phi \quad (1)$$

Onde:

τ = resistência ao cisalhamento

c = coesão

σ = tensão normal

ϕ = ângulo de atrito

Para os diferentes tipos de falhas por cisalhamento, o maciço rochoso pode ser tratado como um material que obedece ao critério de Mohr–Coulomb, no qual a resistência ao cisalhamento é definida em função da coesão (c) e do ângulo de atrito (ϕ) (WYLLIE; MAH, 2004).

4.2.1 Coesão

Conforme exposto por Murrieta (2018), mesmo na ausência de esforços externos aplicados, determinados tipos de solo apresentam uma parcela de resistência ao cisalhamento, denominada por Coulomb como coesão, considerada uma constante do solo.

A resistência ao cisalhamento dos solos decorre, predominantemente, do atrito existente entre as partículas. No entanto, forças de atração química entre essas partículas podem gerar uma parcela de resistência independente da tensão normal atuante no plano de ruptura, caracterizando a chamada coesão real (PINTO, 2006). Segundo Hoque et al. (2023), a coesão é o grau de cisalhamento do solo em um ponto onde a tensão de compressão é igual a zero. A figura 3 traz valores típicos para coesão em argilas.

Figura 3 - Valores típicos de coesão.

Argilas	Coesão, c (kPa)
Muito mole	< 10
Mole	10 a 25
Média	25 a 50
Rija	50 a 100
Muito rija	100 a 200
Dura	> 200

Fonte: Adaptado de Alonso (1983) apud Priebe et al. (2019).

4.2.2 Ângulo de Atrito

O ângulo de atrito do material rochoso está associado ao tamanho e à forma dos grãos expostos na superfície da fratura. Assim, rochas de granulação fina, como aquelas com elevado teor de mica orientada paralelamente à superfície, como os filitos, tendem a apresentar baixos valores de ângulo de atrito, enquanto rochas de granulação grossa, como o granito, apresentam ângulos de atrito mais elevados (WYLLIE; MAH, 2004). Faixas típicas para ângulos de atrito e suas classificações foram apresentados na figura 4.

Figura 4 - Valores típicos de ângulo de atrito para diferentes tipos de rocha.

Classificação	Faixa de ângulo de atrito	Tipos típicos de rocha
Baixo	20–27°	Xistos (alto teor de mica), folhelho, marga
Médio	27–34°	Arenito, siltito, giz, gnaisse, ardósia
Alto	34–40°	Basalto, granito, calcário, conglomerado

Fonte: Adaptado de Wyllie & Mah (2004).

4.2.3 Razão de Poropressão

Wyllie e Mah (2004) destacam que a principal influência da presença de água em uma descontinuidade é a redução da resistência ao cisalhamento, decorrente da diminuição da tensão normal efetiva atuante sobre a superfície. Essa tensão normal efetiva é definida como a diferença entre o peso do maciço rochoso sobrejacente e a pressão de soerguimento gerada pela água. Dessa forma, o efeito da pressão da água, também chamada poropressão (u), sobre a resistência ao cisalhamento pode ser incorporado à equação de resistência ao cisalhamento:

$$\tau = c + (\sigma - u) \tan \phi \quad (2)$$

A Razão de Poropressão (r_u), poder ser definida pela relação entre poropressão e tensão normal:

$$r_u = \frac{u}{\sigma} \quad (3)$$

Onde:

r_u = razão de poropressão;

u = poropressão

σ = tensão normal

4.3 Peso Específico

Segundo Cintra (2003), na ausência de ensaios laboratoriais, o peso específico do solo pode ser estimado a partir de valores aproximados, definidos em função da consistência das argilas e do grau de compacidade das areias. A consistência dos solos está diretamente relacionada ao índice de resistência à penetração (N_{SPT}), conforme estabelecido pela Associação Brasileira de Normas Técnicas (ABNT), por meio da Norma Brasileira NBR 6484:2001. Nesse contexto, Godoy (1972, apud CINTRA, 2003) propõe valores típicos para a estimativa do peso específico de solos.

Figura 5 - Peso Específico de Solos Argilosos.

N (golpes)	Consistência	Peso específico (kN/m ³)
≤ 2	Muito mole	13
3 – 5	Mole	15
6 – 10	Média	17
11 – 19	Rija	19
≥ 20	Dura	21

Fonte: Godoy (1972, apud CINTRA, 2003).

Figura 6 - Peso Específico de Solos Arenosos.

N (golpes)	Consistência	Peso específico (kN/m ³)		
		Areia seca	Úmida	Saturada
< 5	Fofa	16	18	19
5 – 8	Pouco compacta	16	18	19
9 – 18	Medianamente compacta	17	19	20
19 – 40	Compacta	18	20	21
> 40	Muito compacta	18	20	21

Fonte: Godoy (1972, apud CINTRA, 2003).

4.4 Método do Equilíbrio Limite

O método baseia-se na determinação do equilíbrio de uma massa ativa de solo, a qual pode ser delimitada por uma superfície de ruptura de geometria circular, poligonal ou de outra forma qualquer. Neste método, assume-se que a ruptura ocorre ao longo dessa superfície e que todos os elementos que a compõem atingem simultaneamente a condição correspondente a um fator de segurança (FS) igual a 1 (GERSCOVICH, 2016).

4.4.1 Fator de Segurança

O objetivo da análise de estabilidade consiste em avaliar a probabilidade de ocorrência de movimentos de massa em taludes naturais ou artificiais. Essas análises baseiam-se na comparação entre as tensões cisalhantes mobilizadas e a resistência ao cisalhamento ao longo de superfícies potenciais de ruptura (GERSCOVICH, 2016). Portanto, define-se um Fator de Segurança (FS), que pode ser dado pela relação apresentada na figura 7.

Figura 7 - Relação do Fator de Segurança.

$$FS = \frac{\tau_f}{\tau_{mob}}$$

τ_f τ_{mob}	$> 1 \Rightarrow$ obra estável $= 1 \Rightarrow$ ocorre ruptura $< 1 \Rightarrow$ não tem significado físico
--------------------------	--

Fonte: Gerscovich (2016).

Onde:

FS = fator de segurança;

τ_f = resistência ao cisalhamento do solo

τ_{mob} = tensões cisalhantes mobilizadas

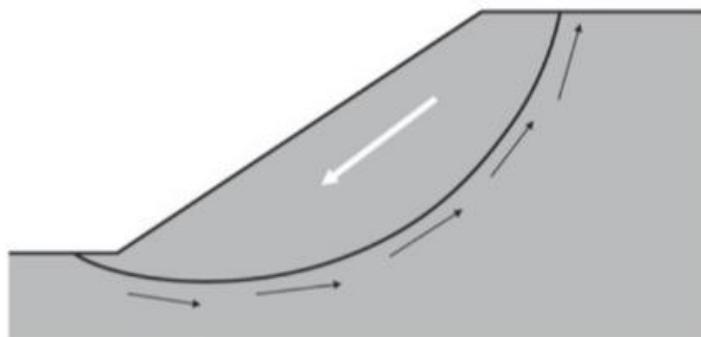
4.5 Ruptura Circular

De maneira geral, a instabilidade de taludes é comumente associada a planos que dividem o maciço em um meio descontínuo. Nesses casos, uma ou mais descontinuidades tendem a controlar a superfície de ruptura, entretanto, em solos ou rochas muito fraturadas ou intemperizadas, o padrão estrutural é pouco definido, permitindo que a superfície de deslizamento se desenvolva ao longo do trajeto de menor resistência no talude, assumindo, na maioria das ocorrências já observadas, uma geometria aproximadamente circular (WYLLIE; MAH, 2004).

Figura 8 - Ruptura Circular.



Ⓐ Vista de ruptura circular (GeoRio, 1999)



Ⓑ Esquema de superfície de ruptura na seção central

Fonte: Gerscovich (2016).

4.6 Estatística Multivariada

De acordo com Hair et al. (2009), a análise multivariada engloba o conjunto de técnicas estatísticas destinadas à avaliação simultânea de múltiplas medidas associadas a indivíduos ou objetos de estudo. Dessa forma, qualquer análise simultânea de mais de duas variáveis pode ser, em princípio, classificada como multivariada.

Abaixo, seguem alguns conceitos estatísticos importantes e as definições propostas por Hair et al. (2009):

- **Homoscedasticidade** – “Quando a variância dos termos de erro (e) parece constante ao longo de um domínio de variáveis preditoras, diz-se que os dados são homoscedásticos. A suposição de variância igual do erro E da população (onde E é estimado a partir de e) é crítica para a aplicação correta de muitas técnicas multivariadas.” (HAIR et al., 2009, p. 51).
- **Linearidade** – “Usada para expressar o conceito de que o modelo possui as propriedades de aditividade e homogeneidade. Em termos gerais, os modelos lineares preveem valores que recaem em uma linha reta que tem uma mudança com unidade constante (coeficiente angular) da variável dependente em relação a uma mudança com unidade constante da variável independente.” (HAIR et al., 2009, p. 51).
- **Normalidade** – “Grau em que a distribuição dos dados da amostra corresponde a uma distribuição normal.” (HAIR et al., 2009, p. 51).
- **Resíduo** – “Parte de uma variável dependente não explicada por uma técnica multivariada. Associado a métodos de dependência que tentam prever a variável dependente, o resíduo representa a parte inexplicada da mesma. Os resíduos podem ser usados em procedimentos diagnósticos para identificar problemas na técnica de estimação ou para identificar relações não especificadas.” (HAIR et al., 2009, p. 51).

4.6.1 Correlação entre variáveis

4.6.1.1 Fórmula do coeficiente de correlação de Pearson

De acordo com Montgomery et al. (2021), o coeficiente de correlação da amostra, r_{xy} , constitui uma medida quantitativa da força da relação linear entre duas variáveis aleatórias, x e y . Esse coeficiente também é conhecido como coeficiente de correlação de Pearson, e é definido através da equação:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{[\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2]^{1/2}} \quad (4)$$

onde:

- x_i e y_i são os valores observados;
- \bar{x} e \bar{y} são as médias das variáveis.

Os valores do coeficiente de Pearson podem variar entre -1 e $+1$, em que valores próximos de ± 1 indicam forte correlação linear, enquanto valores próximos de 0 indicam fraca associação.

4.6.2 Regressão linear múltipla

A regressão múltipla é uma técnica estatística empregada para analisar a relação entre uma única variável dependente (critério) e diversas variáveis independentes (preditoras). Seu objetivo é utilizar variáveis independentes com valores conhecidos para estimar ou prever os valores da variável dependente definida pelo pesquisador (Hair et al., 2009).

4.6.2.1 Forma geral do modelo

Em um modelo de regressão múltipla, a variável dependente (Y) será determinada por mais de uma variável independente (X) (MAIA, s.d.).

Montgomery et al. (2021) sugere a seguinte forma geral do modelo de regressão:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (5)$$

Onde:

β_0 = é o valor esperado de Y quando todos as variáveis independentes forem nulas;

β_1 = coeficiente parcial de regressão para a variável independente X_1 ;

β_2 = coeficiente parcial de regressão para a variável independente X_2 ;

β_k = coeficiente parcial de regressão para a variável independente X_k ;

ε = erro;

O parâmetro β_0 é a interseção do plano. Os coeficientes β_1 e β_2 são denominados coeficientes parciais de regressão, pois β_1 expressa a variação esperada em Y para cada unidade de variação em x_1 , mantendo-se x_2 constante, enquanto β_2 representa a variação esperada em Y para cada unidade de variação em x_2 , mantendo-se x_1 constante (MONTGOMERY et al., 2021).

4.6.3 Teste de esfericidade de Bartlett

Conforme Dutt-Ross (2020), o teste de Bartlett é empregado para verificar a homogeneidade das variâncias entre amostras, ou seja, se elas podem ser consideradas estatisticamente iguais. A verificação desse pressuposto é requerida pela maioria dos procedimentos estatísticos.

De acordo com Bartlett (1951) apud Lugu (2023), o teste de esfericidade de Bartlett, é utilizado para avaliar se a matriz de correlação observada é estatisticamente diferente de uma matriz identidade, na qual as variáveis não apresentam correlação entre si. A hipótese nula do teste pressupõe que os coeficientes de correlação fora da diagonal principal sejam iguais a zero, indicando ausência de interdependência entre as variáveis. Quando o resultado do teste apresenta p-valor inferior a 0,05, rejeita-se essa hipótese, evidenciando a existência de correlações significativas entre as variáveis.

4.6.4 Detecção de *outliers* multivariados

De acordo com Hair et al. (2009), a detecção de outliers multivariados pode ser realizada por meio da medida D^2 de Mahalanobis, que consiste em uma avaliação multivariada de cada observação em relação a um conjunto de variáveis. Ainda segundo os autores, esse método mede a distância de cada observação, em um espaço multidimensional a partir do centro médio de todas as observações, fornecendo um único valor para cada caso, independentemente do número de variáveis analisadas. Valores mais elevados de D^2 indicam observações mais distantes da distribuição geral dos dados nesse espaço multidimensional.

5. MATERIAIS E MÉTODOS

O principal objetivo deste capítulo é expor a metodologia constituída no trabalho, bem como apresentar o banco de dados, *software* e os métodos aplicados durante seu desenvolvimento. Os algoritmos resultantes estão dispostos no Apêndices do presente trabalho.

5.1 Banco de Dados

O banco de dados utilizado neste estudo é oriundo do trabalho “*Prediction of slope stability using Tree Augmented Naive-Bayes classifier: modeling and performance evaluation*” (AHMAD et al, 2022). Ele é composto por 87 elementos amostrais referentes a taludes sujeitos predominantemente a mecanismos de ruptura circular, contendo 7 variáveis que foram utilizadas para construção dos modelos de regressão. O banco de dados pode ser encontrado no Anexo A deste trabalho.

5.1.1 Variáveis do estudo

O conjunto de variáveis analisadas é constituído por parâmetros geotécnicos e geométricos, além do Fator de Segurança. São elas:

- peso específico (γ);
- coesão (c);
- ângulo de atrito interno (ϕ);

- ângulo de inclinação do talude (β);
- altura do talude (H);
- razão de poropressão (r_u);
- fator de segurança (FS);

O FS foi considerado como variável dependente, enquanto as demais foram tratadas como variáveis independentes do modelo.

5.2 Ambiente computacional

As análises estatísticas e a construção dos modelos de regressão foram realizadas no ambiente RStudio, utilizando a linguagem R. Foram utilizados *scripts* para leitura, tratamento, análise exploratória e modelagem dos dados, bem como pacotes específicos para aplicação de testes estatísticos e análise multivariada.

5.2.1 RStudio

O RStudio consiste em um ambiente integrado de desenvolvimento (Integrated Development Environment – IDE) desenvolvido para dar suporte a várias linguagens de programação, com destaque para a linguagem R.

5.3 Análise exploratória dos dados

Foi realizada uma análise exploratória do banco de dados com o objetivo de compreender o comportamento das variáveis e suas relações. Nessa etapa, foram efetuados cálculos de estatística básica, incluindo valores mínimos e máximos, mediana e média aritmética. Além disso, foram calculadas as matrizes de covariâncias e de correlação, a fim de identificar a dependência linear entre as variáveis.

Foram construídos *boxplots*, histogramas e gráficos de dispersão, com o intuito de identificar previamente se as variáveis apresentavam comportamento de distribuição normal, assimetrias, possíveis *outliers* e tendências nos dados.

Uma matriz de dispersão foi obtida por meio da utilização da função *pairs()* do software RStudio, permitindo avaliar, de forma bivariada e visual, as correlações entre as variáveis. De maneira complementar, a função *painel.pearson* exibe, no painel

superior da matriz de dispersão, o valor absoluto do coeficiente de correlação linear entre os pares. Esse coeficiente pode variar entre 0 e 1, em que valores abaixo de 0,3 indicam associação linear fraca, enquanto valores próximos de 1 indicam forte associação linear.

5.4 Testes estatísticos preliminares

A fim de verificar se o banco de dados era adequado para aplicação de técnicas multivariadas, foram realizados testes estatísticos preliminares capazes de avaliar a existência de correlações significativas, bem como a viabilidade do uso dos métodos aplicados.

5.4.1 Teste de Esfericidade de Bartlett

Neste teste, a hipótese nula considera a inexistência de correlação significativa entre as variáveis. Para satisfazer o interesse desse estudo, assume-se o que é estabelecido previamente na literatura, segundo a qual a hipótese nula deve ser rejeitada para um *p-value* inferior a 0,05, indicando, com 95% de confiança, a presença de correlação suficiente para a aplicação de métodos multivariados.

Também foi realizado um teste de normalidade multivariada, visando avaliar, de forma simultânea, se todas as variáveis do conjunto de dados seguem uma distribuição normal multivariada, e não apenas as suas distribuições individuais. Para realização deste teste foi utilizada a função *mvn()* do pacote MVN do RStudio.

5.5 Identificação e tratamento de *outliers*

A identificação de *outliers* foi realizada por meio da distância de Mahalanobis, técnica comumente utilizada em análises multivariadas para encontrar elementos amostrais discrepantes em relação ao conjunto de dados. As observações identificadas como *outliers* foram analisadas e removidas conforme os critérios definidos em cada cenário proposto.

5.6 Construção dos modelos de regressão

Foram construídos modelos de regressão linear múltipla considerando três cenários distintos:

- **Análise 1:** construção do modelo utilizando todos os dados, sem tratamento de dados e sem remoção de *outliers*;
- **Análise 2:** construção do modelo utilizando todo o conjunto de dados, mas com a remoção dos *outliers* identificados pela distância de Mahalanobis;
- **Análise 3:** tratamento prévio dos dados baseado em critérios de valores típicos e observações para algumas variáveis independentes, seguido da remoção de outliers.

Como dito anteriormente, o Fator de Segurança (FS) foi considerado a variável dependente, sendo as demais variáveis utilizadas como variáveis independentes.

Os modelos foram ajustados utilizando a função *lm()* do R, que estima os coeficientes de determinação por meio do método dos mínimos quadrados ordinários.

5.6.1 Critérios adotados para tratamento dos dados na Análise 3

Para a análise 3 alguns critérios adicionais foram implementados. Foram removidos os elementos amostrais que apresentaram pelo menos uma das seguintes condições:

- Peso específico (γ) < 15kN/m³
- Coesão (c) < 5 kPa
- Altura do talude (H) > 50 m

Esses limites foram definidos com base em valores da literatura, propostos para peso específico por Godoy (1972, apud Cintra, 2003) e para coesão por Alonso (1983). Além da análise do próprio banco de dados para definir o critério adotado para a variável altura, buscando calibrar o modelo com condições que representassem o comportamento médio dos taludes. Importante ressaltar que tais critérios foram adotados por representarem condições típicas de taludes artificiais e encostas em solos residuais, permitindo que o modelo estatístico fosse ajustado a situações geomecânicas compatíveis a cenários recorrentes em projetos.

5.7 Análise dos modelos

Os modelos obtidos foram analisados por meio do resultado referente ao coeficiente de determinação múltiplo (R^2) e do coeficiente de determinação ajustado (R^2 ajustado), que permitem avaliar a capacidade explicativa dos modelos considerando o número de variáveis independentes e quais variáveis apresentavam maior significância na análise.

Além disso, foi realizada a análise dos resíduos com o objetivo de verificar a normalidade e a homoscedasticidade dos mesmos, garantindo que eles não apresentem qualquer tipo de padrão ou tendência, confirmando que a variabilidade dos erros é aleatória.

5.8 Validação dos modelos

Como etapa adicional, foi implementada a validação cruzada do tipo *K-fold*, com valores de *K* iguais a 5 e 10. Ela permitiu avaliar a capacidade de generalização e o desempenho preditivo dos modelos. O conjunto de dados foi dividido aleatoriamente em *K* subconjuntos, de modo que, a cada iteração, um *fold* fosse utilizado como conjunto de teste e os demais como conjunto de treinamento.

Em cada iteração, o modelo foi ajustado exclusivamente com os dados de treino e aplicado ao conjunto de teste para a predição do fator de segurança. O desempenho foi avaliado por meio do erro quadrático médio da raiz (RMSE) e do coeficiente de determinação preditivo (R^2). Ao final, as médias foram utilizadas para identificar se o modelo possui boa capacidade de generalização.

6. RESULTADOS E DISCUSSÕES

Este capítulo apresenta os resultados obtidos a partir da aplicação de modelos de regressão linear múltipla para estimativa do Fator de Segurança (FS) em taludes. A comparação entre os cenários permitiu avaliar a influência e a importância do pré-processamento para o melhor desempenho dos modelos. Cada cenário foi analisado individualmente por meio de estatística descritiva, análise gráfica e ajuste do modelo de regressão.

6.1 Análise Preliminar

Inicialmente, foi realizada uma análise preliminar, a fim de entender o banco de dados e verificar a consistência dos mesmos. Essa etapa contribuiu para a identificação de valores física ou geotecnicamente incoerentes com a análise proposta. Esses valores geralmente podem ser associados a erros de digitação, conversão de unidades ou falhas na construção do banco de dados.

Como essas observações poderiam prejudicar o ajuste estatístico dos modelos de regressão, bem como a análise dos resultados, elas foram removidas. Assim, o banco de dados inicial que antes continha 87 amostras, foi reduzido para 71, garantindo maior coerência e confiabilidade para as etapas seguintes.

6.2 Cenários de Análise

Para avaliar a influência do tratamento dos dados no desempenho e na confiabilidade dos modelos de regressão linear, foram desenvolvidos três cenários de modelagem, que se diferenciam com base no pré-processamento feito em cada um deles, conforme já exposto no Capítulo 5, Seção 5.6 - Construção dos Modelos de Regressão.

6.3 Teste de Bartlett

Como pode ser observado na tabela 1, em todos os cenários analisados o teste de Bartlett resultou em um *p-value* menor que 0,05. Portanto, a hipótese nula pode ser descartada, indicando correlações significativas com confiança superior a 95%.

Tabela 1 - Resultados Teste de Bartlett.

Análise	<i>p-value</i>
1	3.735362e-25
2	
3	1.431403e-11

Fonte: Autor (2026).

6.4 Análise 1

6.4.1 Estatística descritiva

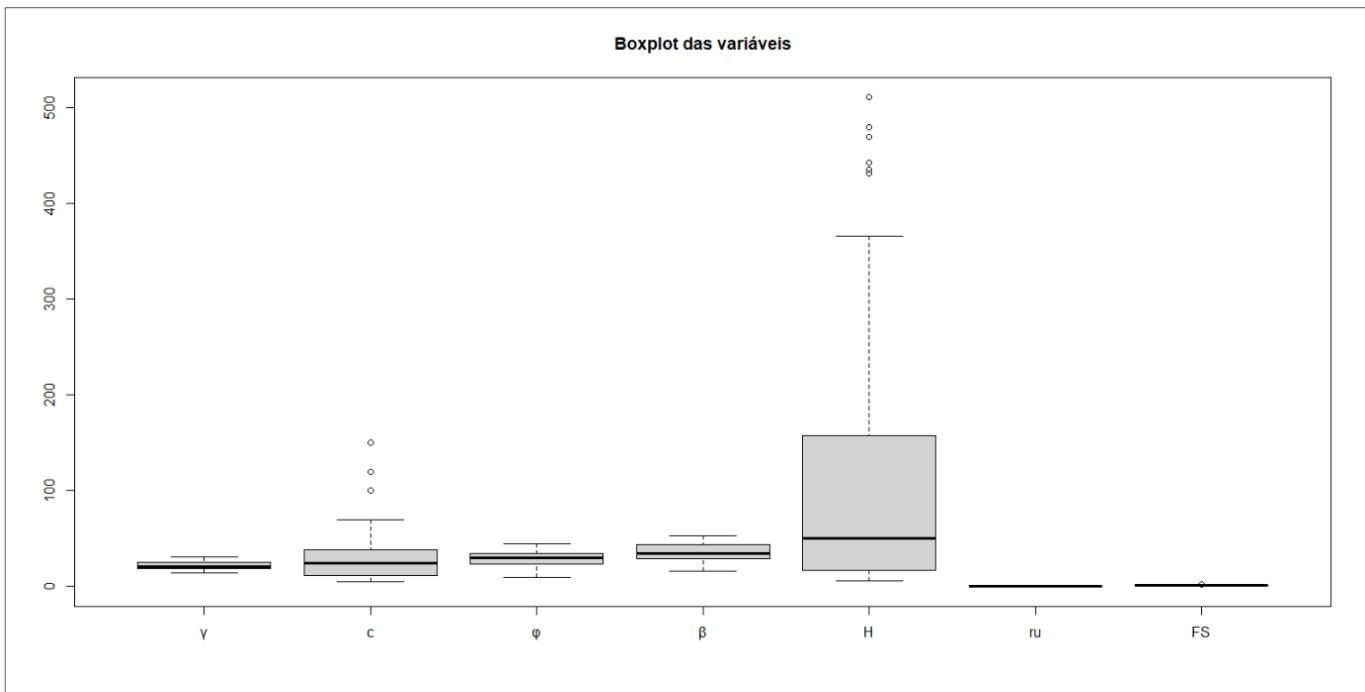
A análise inicial foi realizada utilizando o banco de dados após a etapa de análise preliminar e remoção de dados incoerentes, porém preservando as observações originais sem avaliação estatística para remoção de dados discrepantes. A Tabela 2 apresenta o resumo da estatística descritiva das variáveis consideradas.

Tabela 2 - Resumo Estatística Descritiva – Análise 1.

Dados	γ	c	φ	β	H	ru	FS
Min.	14	5	10	16	6	0	0.625
1º Qu.	18.8	12	23.5	29.6	17.5	0.1	1.11
Mediana	21	24.8	30	35	50	0.25	1.25
Média	21.83	30.05	28.91	35.28	118.7	0.2265	1.321
3º Qu.	25	38.35	35	44.25	157.5	0.3	1.455
Max.	31.3	150	45	53	511	0.5	2.31

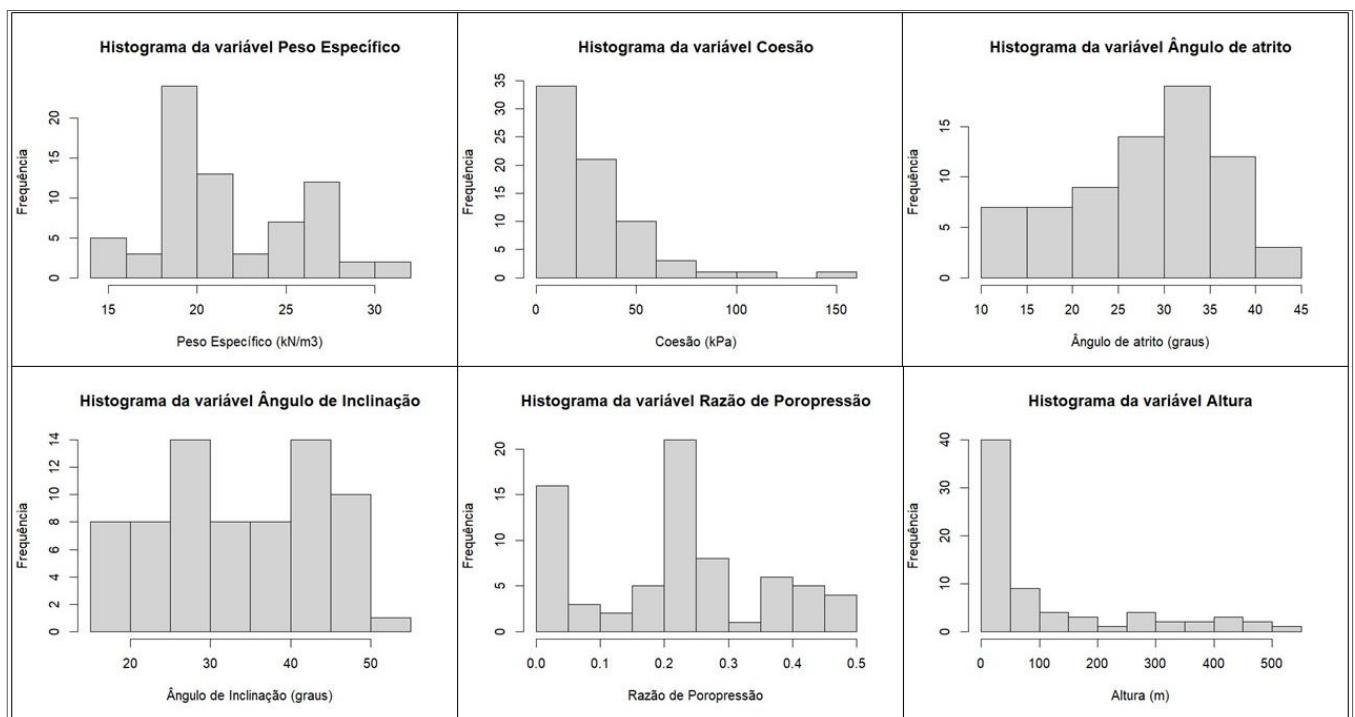
Fonte: Autor (2026).

Os resultados desta etapa mostram certo nível de dispersão dos parâmetros analisados, especialmente para as variáveis coesão (c) e altura (H), cujos valores máximos são muito superiores às medidas centrais. Esse comportamento pode ser observado também no *boxplot* apresentado na Figura 9.

Figura 9 - Boxplot Geral – Análise 1

Fonte: Autor (2026).

A presença de *outliers* univariados coloca em evidência o quanto heterogêneo é este conjunto de dados, característica que pode ser comum devido às diferentes condições em taludes observados nas análises em campo.

Figura 10 - Histogramas - Análise 1.

Fonte: Autor (2026).

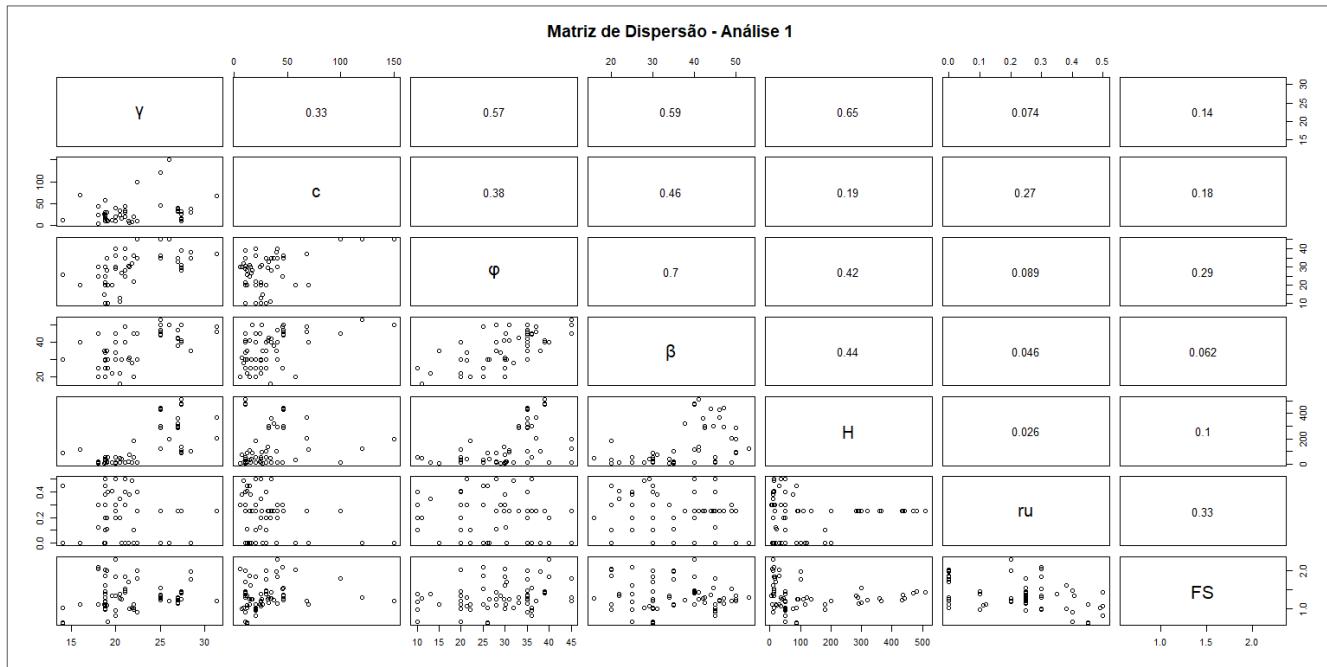
Na Figura 10, os histogramas apresentados indicam assimetrias e distribuições não normais para diversas variáveis, confirmando, novamente, a variabilidade natural dos parâmetros.

Do ponto de vista estatístico, dados muito dispersos e com valores extremos podem comprometer alguns pressupostos da regressão linear, como a homocedasticidade e a estabilidade das estimativas. Já em termos geotécnicos, essa variação representa a heterogeneidade natural dos maciços de solo, mas dificulta a construção de modelos com boa capacidade de previsão.

6.4.2 Correlações entre as variáveis

A matriz de dispersão apresentada na Figura 11 permitiu avaliar visualmente as relações bivariadas entre as variáveis do modelo.

Figura 11 - Matriz de dispersão – Análise 1.



Fonte: Autor (2026).

Observa-se alguns coeficientes de correlação abaixo de 0,3 entre o FS e grande parte das variáveis independentes, além de alta dispersão dos pontos. Esse comportamento indica baixa linearidade entre as variáveis, o que tende a comprometer o desempenho do modelo de regressão.

6.4.3 Desempenho do modelo

Para avaliar se o ajuste do modelo foi satisfatório, espera-se que os valores médios dos resíduos tendam a zero e que não apresentem nenhum tipo de tendência ou normalidade. Como pode ser observado na tabela 3, o valor de medida central dos erros é igual a 0.0094, dando um primeiro indício de que o ajuste do modelo foi bom.

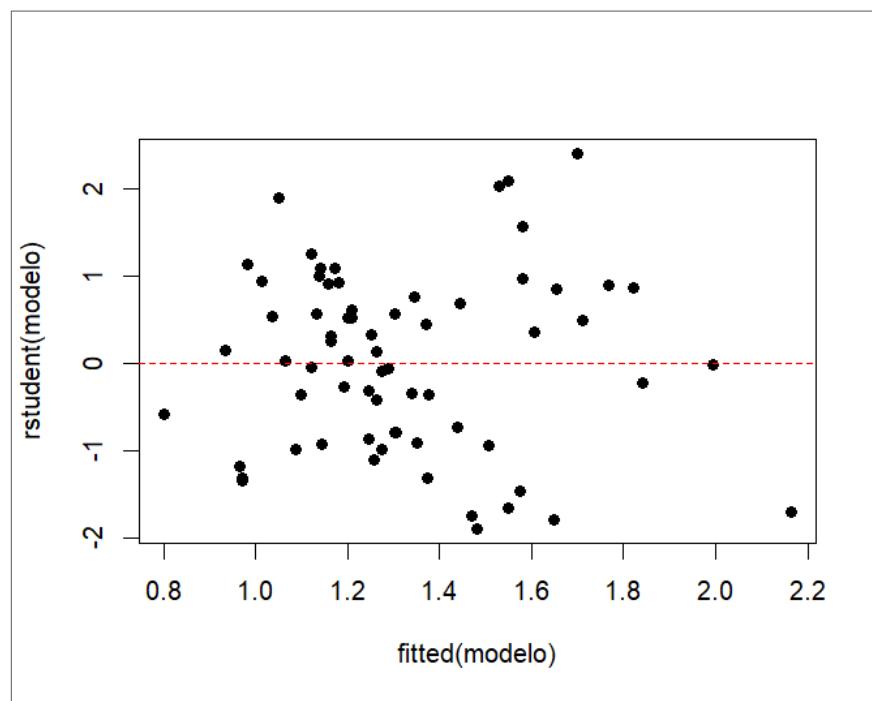
Tabela 3 - Resíduos – Análise 1.

Min	1Q	Mediana	3Q	Max
-0.585	-0.2705	0.0094	0.2004	0.7478

Fonte: Autor (2026).

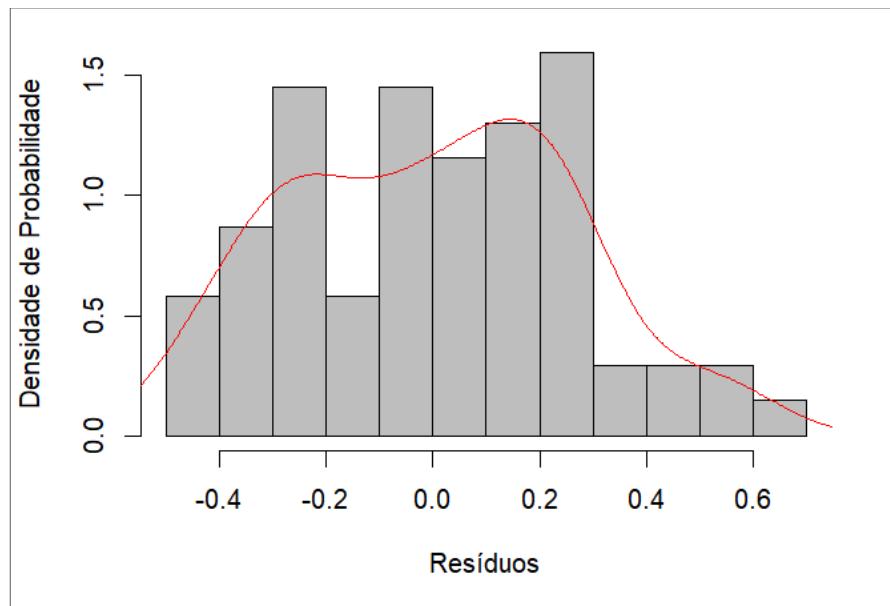
Nos gráficos presentes nas figuras 12 e 13 podemos avaliar a dispersão e a normalidade dos resíduos, podendo confirmar o ajuste satisfatório do modelo.

Figura 12 - Dispersão dos resíduos – Análise 1.



Fonte: Autor (2026).

Figura 13 - Histograma dos resíduos – Análise 1.



Fonte: Autor (2026).

6.4.3.1 Coeficientes de Regressão

Na tabela 4 é possível observar os coeficientes resultantes da regressão para cada um dos parâmetros utilizados para calibração do modelo.

Tabela 4 - Coeficientes de Regressão – Análise 1.

Parâmetro	Coeficiente Estimado	Significância
Constante	0.9864411	***
γ	0.0261668	.
c	0.0017277	
ϕ	0.0242515	***
β	-0.0214203	***
H	-0.0008062	*
ru	-0.6102048	*

Fonte: Autor (2026).

Na tabela, é possível observar que os parâmetros ângulo de atrito (ϕ) e ângulo de inclinação do talude (β) apresentaram o maior nível de significância (***) , indicando forte relevância estatística para o ajuste do modelo. Geotecnicalemente, isso é

coerente, pois o ângulo de atrito está diretamente associado à resistência ao cisalhamento do solo como visto anteriormente na equação 1, que representa o critério de resistência de Mohr-Coulomb.

A variável altura do talude (H) e a razão de poropressão (ru) apresentaram significância intermediária (*), indicando que exercem influência, ainda que menor. A poropressão reduz as tensões efetivas, diminuindo a resistência ao cisalhamento, conforme dito por Wyllie e Mah (2004).

A coesão (c) apresentou baixa significância estatística neste cenário, isso pode estar associado à elevada variabilidade dessa variável no domínio analisado, reduzindo sua capacidade de explicar linearmente o comportamento do FS.

6.4.3.2 Fórmula do modelo de regressão

A equação 6 é a equação resultante do modelo ajustado para o cenário 1, baseado na forma geral dos modelos de regressão linear múltipla.

$$FS = 0.9864411 + 0.0261668\gamma + 0.0017277c + 0.0242515\phi - 0.0214203\beta - 0.0008062H - 0.6102048ru \quad (6)$$

Por fim, para avaliar se o modelo apresentou bom desempenho, se verifica o coeficiente de determinação ajustado (R^2 ajustado). Neste cenário ele foi igual a 0,3193, indicando baixo poder explicativo. Esse resultado evidencia que o modelo consegue explicar apenas uma parcela limitada dos dados originais.

6.4.4 Avaliação do modelo

Mesmo com após a realização da análise preliminar, o modelo apresentou baixo desempenho. Isso se dá devido à alta variabilidade dos dados e a presença de *outliers*, que prejudicam o ajuste estatístico do modelo e a relação linear entre as variáveis.

Do ponto de vista geotécnico, essa alta variabilidade engloba diferentes contextos físicos de taludes, mas estatisticamente reduz a consistência do modelo, o tornando menos eficaz, mostrando a limitação do uso direto de dados brutos em modelos multivariados.

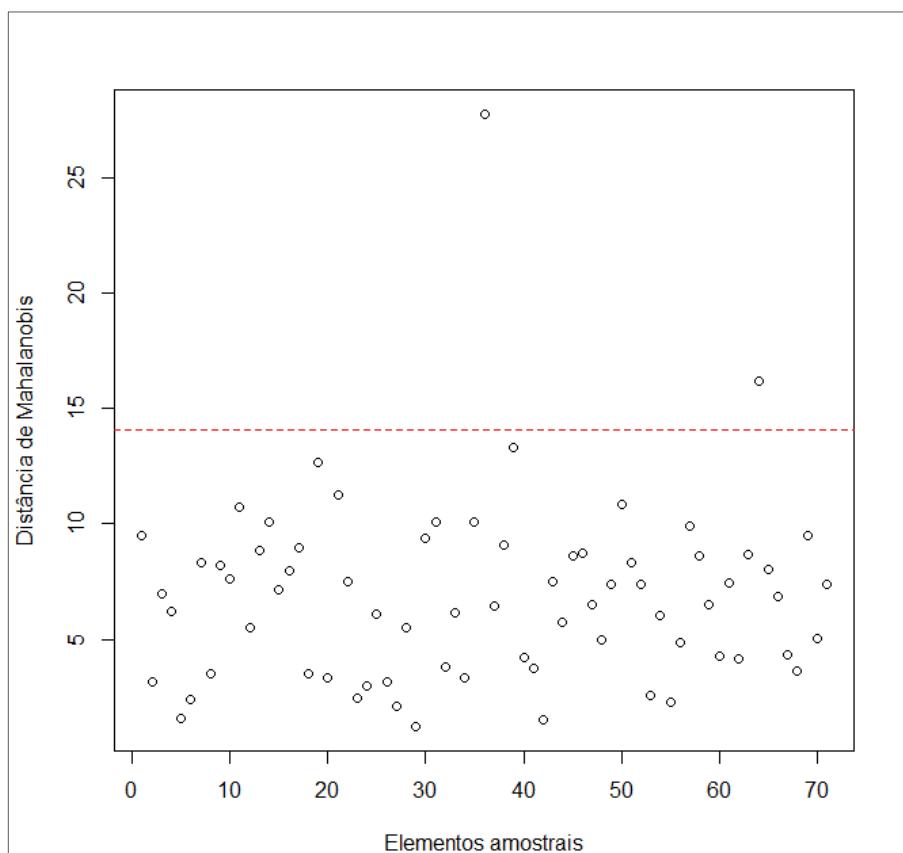
6.5 Análise 2

Considerando que a Análise 2 foi realizada com base no mesmo banco de dados da Análise 1, admite-se que as conclusões inciais referentes à estatística descritiva permanecem as mesmas da etapa anterior. A principal diferença entre as análises decorre da remoção de *outliers*, buscando o aumento da confiabilidade do modelo.

6.5.1 Identificação de *outliers*

A detecção foi realizada com base na distância de Mahalanobis, cujo limite crítico determinado foi igual a 14,067. A partir desse critério, foram identificados dois elementos amostrais classificados como *outliers* multivariados, correspondentes às observações 41 e 78.

Figura 14 - Distâncias de Mahalanobis e limite crítico – Análise 2.



Fonte: Autor (2026).

6.5.2 Estatística Descritiva pós remoção de outliers

Na Análise 2, foram removidos dois elementos amostrais identificados como *outliers* multivariados. A Tabela 5 apresenta o resumo da estatística descritiva após essa remoção.

Tabela 5 - Resumo Estatística Descritiva pós remoção de outliers – Análise 2.

Dados	γ	c	ϕ	β	H	ru	FS
Min.	14	5	10	16	6	0	0.625
1º Qu.	18.8	12	22	29.2	15	0.11	1.11
Mediana	20.6	24	30	35	50	0.25	1.25
Média	21.72	27.01	28.44	34.81	117.5	0.2331	1.323
3º Qu.	25	35	35	42.6	135	0.3	1.46
Max.	31.3	100	45	50	511	0.5	2.31

Fonte: Autor (2026).

Observa-se leve redução dos valores máximos e da dispersão das variáveis, especialmente para a coesão, indicando que os dados estão pouco mais homogêneos em comparação à Análise 1.

6.5.3 Desempenho do modelo

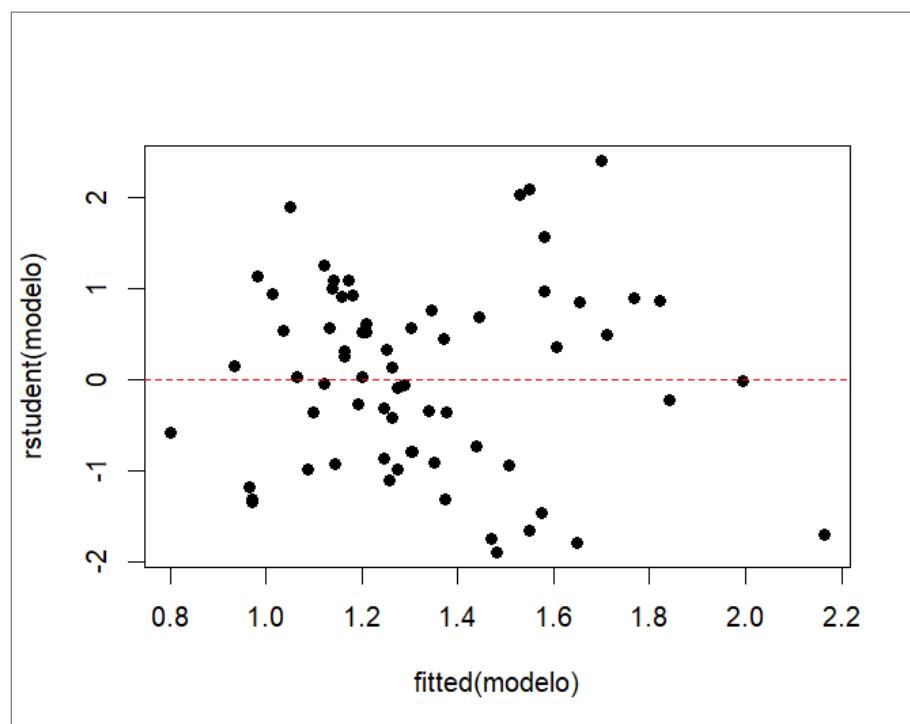
Neste segundo cenário, a análise inicial dos resíduos sugere que o ajuste do modelo teve bom nível de satisfação, conforme dados dispostos na tabela 6. Também é possível fazer essa conclusão a partir do comportamento observado nos gráficos das figuras 15 e 16.

Tabela 6 - Resíduos – Análise 2.

Min	1Q	Mediana	3Q	Max
-0.49117	-0.24031	0.00634	0.20501	0.61041

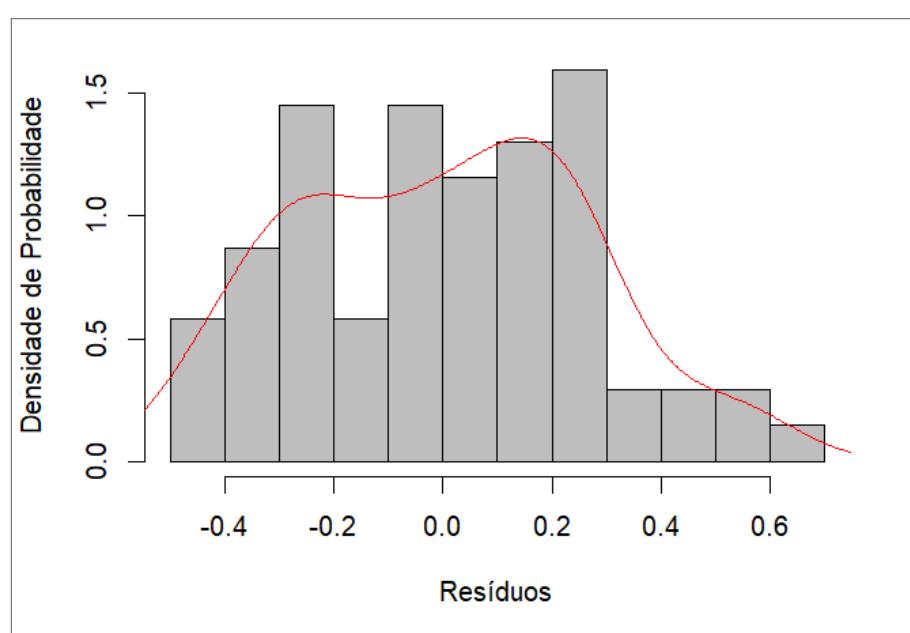
Fonte: Autor (2026).

Figura 15 - Dispersão dos Resíduos – Análise 2.



Fonte: Autor (2026).

Figura 16 - Histograma dos resíduos – Análise 2.



Fonte: Autor (2026).

6.5.3.1 Coeficientes de regressão

Na tabela 7 estão dispostos os coeficientes de regressão resultantes deste segundo cenário.

Tabela 7 - Coeficiente de Regressão – Análise 2.

Parâmetro	Coeficiente Estimado	Significância
Constante	0.9308533	***
γ	0.0219366	.
c	0.0071211	**
ϕ	0.0282661	***
β	-0.0235557	***
H	-0.0008769	**
ru	-0.6759952	**

Fonte: Autor (2026).

Após a remoção dos *outliers*, observa-se aumento da significância da coesão (c) (**), indicando que com a redução da influência dos dados discrepantes, o modelo passa a entender de forma mais clara o seu papel na estimativa do FS.

O ângulo de atrito (ϕ) e o ângulo do talude (β) permanecem significativos (***) , confirmando sua importância geométrica no equilíbrio das forças. Já a altura (H) elevou seu grau de significância (**), reforçando a ideia de que o aumento de massa potencialmente instável proporcionado pela variação deste parâmetro, influencia negativamente o FS, indo de encontro a afirmação feita por Santos (2016).

A razão de poropressão (ru) ganha maior significância neste cenário (**), o que é coerente com seu efeito direto na redução da resistência efetiva.

Portanto, a melhoria geral dos níveis de significância demonstra que a remoção dos *outliers* permitiu ao modelo representar de forma mais fiel os mecanismos físicos envolvidos.

6.5.3.2 Fórmula do modelo de regressão

A equação 7 é resultante deste modelo de regressão:

$$FS = 0.9308533 + 0.0219366\gamma + 0.0071211c + 0.0282661\phi - 0.0235557\beta - 0.0008769H - 0.6759952ru \quad (7)$$

O modelo ajustado após a remoção dos *outliers* apresentou R^2 ajustado igual a 0,4485, representando melhora significativa em relação ao cenário anterior. Esse aumento demonstra a influência negativa que observações discrepantes exerciam sobre o ajuste do modelo.

6.5.4 Avaliação do modelo

A remoção dos *outliers* multivariados contribuiu para o aumento da confiabilidade do modelo, reduzindo desvios estatísticos associadas a estas observações. Portanto, confirma o comportamento esperado em regressões lineares, uma vez que estes elementos exercem grande influência sobre os coeficientes estimados.

Apesar disso, o modelo ainda apresenta desempenho moderado, indicando que a variabilidade característica dos parâmetros geotécnicos permanece relevante mesmo após a exclusão dos casos extremos.

6.6 Análise 3

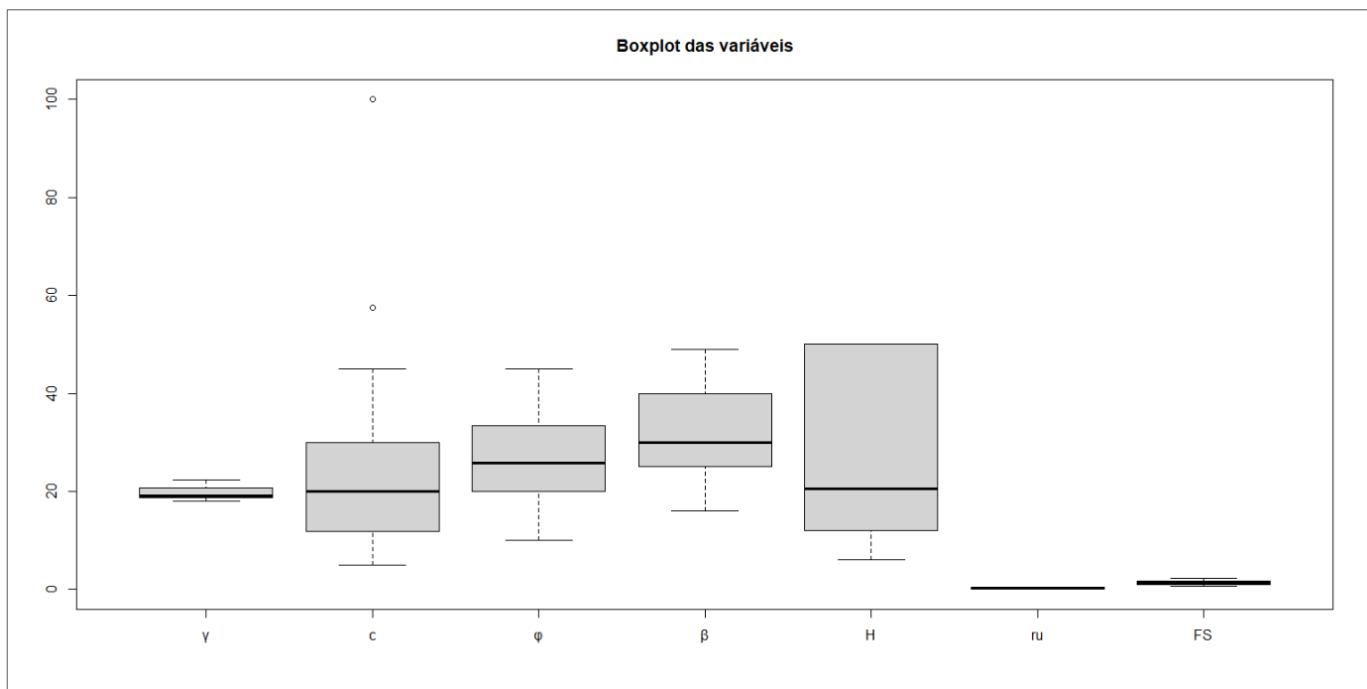
6.6.1 Estatística descritiva

A Tabela 8 apresenta o resumo da estatística descritiva do banco de dados tratado por critérios de valores típicos.

Tabela 8 - Resumo Estatística Descritiva - Análise 3.

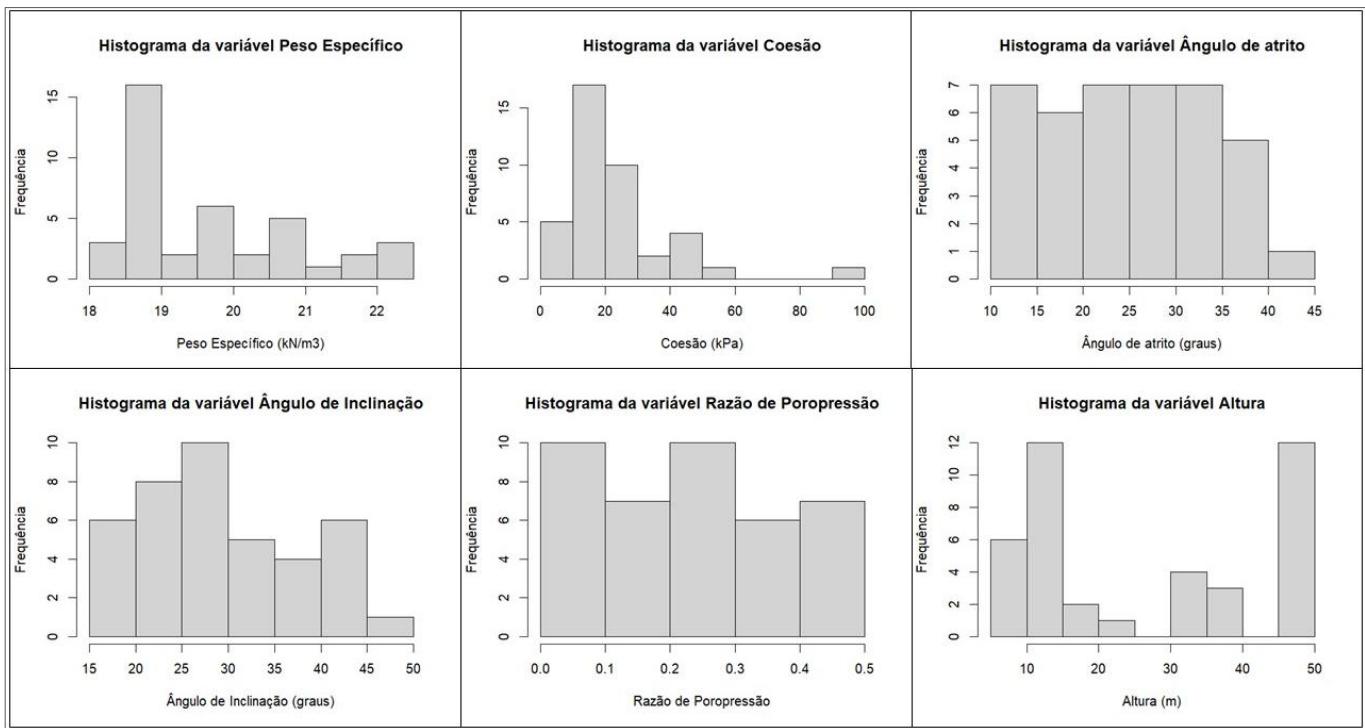
Dados	γ	c	ϕ	β	H	ru	FS
Min.	18	5	10	16	6	0	0.65
1º Qu.	18.8	11.93	20	25	12	0.1075	1.085
Mediana	19.1	20	25.75	30	20.5	0.3	1.315
Média	19.77	24.23	25.84	31.48	27.36	0.2489	1.385
3º Qu.	20.7	30	32.75	40	50	0.4	1.725
Max.	22.4	100	45	49	50	0.5	2.31

Fonte: Autor (2026).

Figura 17 - Boxplot Geral – Análise 3.

Fonte: Autor (2026).

Figura 18 - Histogramas - Análise 3.



Fonte: Autor (2026).

Quando comparado com as análises 1 e 2, existe uma redução significativa da dispersão e maior concentração dos dados em torno das medidas centrais. Isto sugere um conjunto de dados mais homogêneo e representativo das condições médias de estabilidade.

Os dados mais homogêneos tendem a apresentar estimativas mais estáveis e com menor variância, favorecendo o desempenho de modelos lineares e melhora a confiabilidade dos coeficientes estimados.

6.6.2 Correlações entre as variáveis

Figura 19 - Matriz de dispersão – Análise 3.



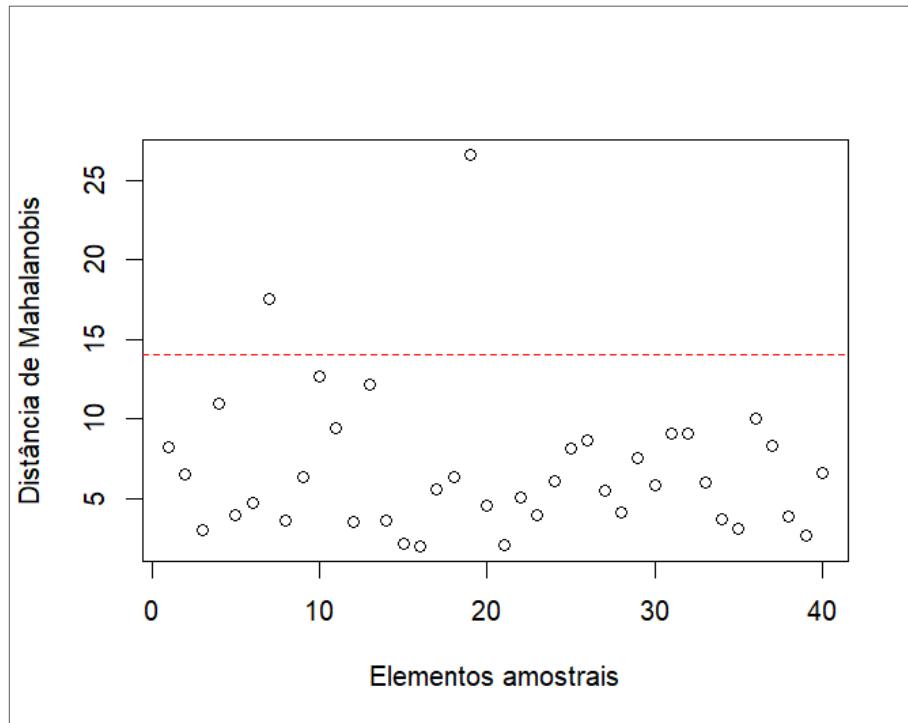
Fonte: Autor (2026).

A matriz evidencia correlações lineares mais fortes entre o Fator de Segurança e as demais variáveis, com menor espalhamento dos pontos, indicando melhor adequação ao modelo.

6.6.3 Identificação de *outliers*

Mais uma vez a distância de Mahalanobis foi utilizada para identificação dos dados discrepantes. O limite crítico calculado foi igual a 14,067. Foram identificados dois elementos amostrais, 17 e 45, classificados como *outliers*.

Figura 20 - Distâncias de Mahalanobis e limite crítico - Análise 3.



Fonte: Autor (2026).

6.6.4 Desempenho do modelo

A análise dos resíduos foi realizada a partir dos dados dispostos na tabela 9 e nas representações gráficas apresentadas a seguir.

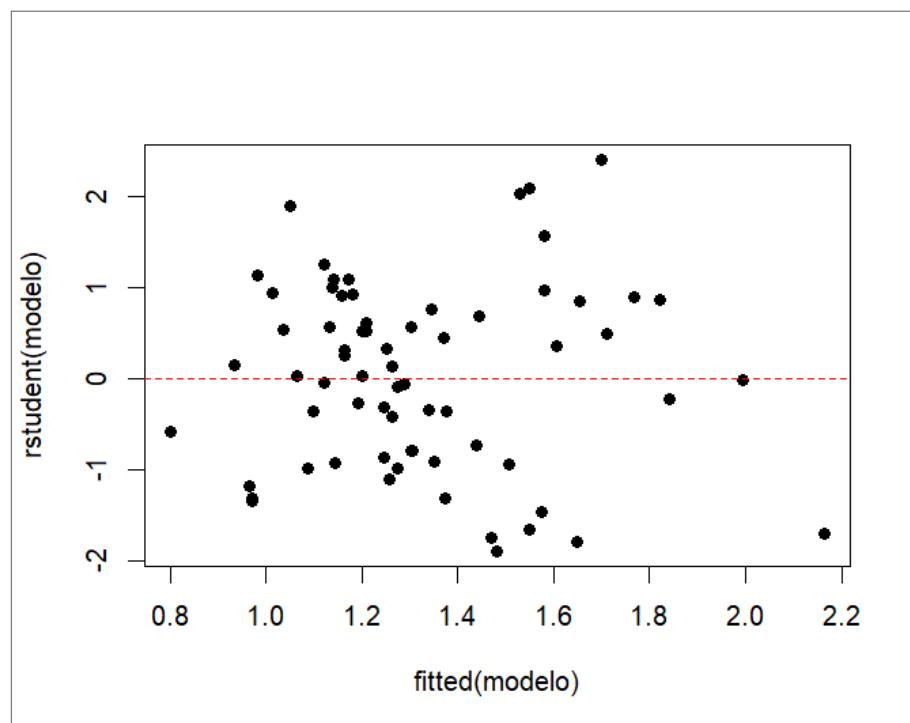
Tabela 9 - Resíduos – Análise 3.

Min	1Q	Mediana	3Q	Max
-0.412	-0.0985	0.00691	0.10647	0.34232

Fonte: Autor (2026).

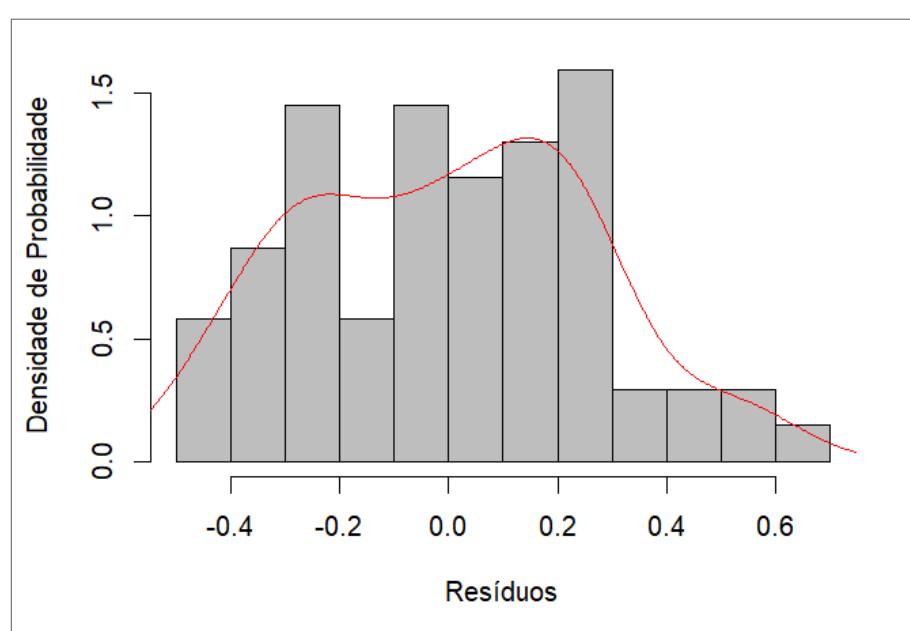
Na figura 21 verifica-se que os resíduos se distribuem de forma aleatória, sem a presença de quaisquer tendências ou padrões, o que indica adequação do ajuste linear. Já na figura 22, é possível ressaltar que, conforme esperado, os resíduos não apresentam normalidade.

Figura 21 - Dispersão dos Resíduos – Análise 3.



Fonte: Autor (2026).

Figura 22 - Histograma dos Resíduos – Análise 3.



Fonte: Autor (2026).

6.6.4.1 Coeficientes de regressão

A tabela 10 apresenta os valores dos coeficientes obtidos pelo modelo de regressão.

Tabela 10 - Coeficiente de Regressão – Análise 3.

Parâmetro	Coeficiente Estimado	Significância
Constante	2.808259	***
γ	-0.050297	.
c	0.015013	***
ϕ	0.019626	**
β	-0.021113	***
H	-0.012561	***
ru	-1.021841	***

Fonte: Autor (2026).

Neste cenário, observa-se que praticamente todas as variáveis apresentam alta significância (**, ***), indicando um modelo estatisticamente consistente e geomecanicamente coerente.

A coesão (c) e o ângulo de atrito (ϕ) apresentam elevada significância, refletindo sua função direta como parâmetros de resistência ao cisalhamento. A inclinação do talude (β) e a altura (H) também são altamente significativas, pois controlam a diretamente o impacto resultante das forças mobilizadoras.

A razão de poropressão (ru) também apresenta forte significância (***) , sendo um dos parâmetros mais sensíveis do modelo. Isso se dá devido ao fato de que qualquer aumento deste parâmetro pode reduzir drasticamente as tensões efetivas, impactando o FS.

O aumento geral da significância estatística nesse cenário indica que o tratamento dos dados reduziu substancialmente as inconsistências, permitindo que o modelo representasse de forma mais clara as relações físicas conhecidas na mecânica dos solos.

6.6.4.2 Fórmula do modelo de regressão

A partir dos coeficientes estimados pelo modelo de regressão linear múltipla, novamente foi deduzida a equação que relaciona o Fator de Segurança às variáveis consideradas, válida apenas para o domínio de dados tratado para este cenário. A equação é apresentada abaixo:

$$FS = 2.808259 - 0.050297\gamma + 0.015013c + 0.019626\phi - 0.021113\beta - 0.012561H - 1.021841ru \quad (8)$$

O modelo ajustado apresentou R^2 ajustado igual a 0,8134, mostrando seu elevado poder explicativo, com desempenho significativamente superior aos cenários analisados anteriormente.

6.6.5 Validação Cruzada

A Tabela 11 apresenta o desempenho do modelo avaliado por validação cruzada K-fold, com $K=5$ e $K=10$, permitindo analisar sua capacidade preditiva em partições independentes de dados.

Tabela 11 - Resultados da Validação Cruzada.

K	RMSE Médio	Desvio Padrão RMSE	R2 Médio	Desvio Padrão R2
5	0.220	0.065	0.781	0.233
10	0.206	0.114	0.744	0.345

Fonte: Autor (2026).

Observa-se que os valores de RMSE médio são baixos em ambos os casos, indicando que o erro médio de predição do Fator de Segurança é baixo quando o modelo é aplicado para dados não utilizados no ajuste.

Já em relação ao coeficiente de determinação, os valores de R^2 médio confirmam que o modelo mantém elevada capacidade explicativa mesmo fora da partição de ajuste, reforçando sua robustez.

Os desvios padrão associados a ambos os índices indicam a estabilidade do modelo entre as diferentes partições dos dados. Para K=5, observa-se menor variabilidade, enquanto em K=10 existe um aumento do desvio padrão, o que é esperado dado o menor tamanho das partições de teste. Ainda assim, os valores permanecem satisfatórios e com um modelo estatisticamente confiável.

6.6.6 Avaliação do modelo

Os resultados obtidos pelo modelo de regressão confirmam que o tratamento dos dados mais robusto resultou em um modelo mais confiável e com maior capacidade de explicativa, indicando que grande parte da variabilidade do FS foi explicada pelo modelo. Já os resultados da validação cruzada confirmam que o modelo apresenta boa capacidade de generalização, não estando restrito apenas ao conjunto de dados utilizado no ajuste.

Entretanto, essa melhoria ocorre à custa de maiores restrições do domínio de validade, limitando sua aplicação a condições geomecânicas semelhantes às consideradas para calibração do modelo. Apesar da redução da variabilidade geral dos dados, foi mantida uma variabilidade compatível com o contexto de solos residuais, preservando a representatividade das condições médias desses materiais. Dessa forma, o modelo torna-se particularmente adequado para aplicação em taludes artificiais e encostas associadas a contextos urbanos, rodoviários e ferroviários, os quais comumente se enquadram nas restrições impostas durante o tratamento dos dados.

Dados os resultados, este modelo foi considerado como o mais robusto entre os cenários de avaliação.

6.7 Síntese das Análises

A tabela 12 traz um resumo dos três cenários analisados.

Tabela 12 - Comparaçāo de resultados dos modelos.

Análise	Tratamento dos Dados	R ² ajustado	Interpretação
1	análise preliminar e modelo calibrado com os dados brutos	0,3193	Baixa representatividade estatística
2	Mesmo banco de dados da análise 1, porém com a remoção de outliers multivariados	0,4485	Melhoria moderada
3	Tratamento dos dados mais robusto, aumentando a restrição do modelo, mas com maior confiabilidade	0,8134	Alta representatividade dos dados, mas com restrições

Fonte: Autor (2026).

Embora os cenários 1 e 2 mantenham maior variabilidade do conjunto de dados, o cenário 3 possui desempenho estatístico superior, tornando-se a alternativa mais adequada para estimativas que representem as condições médias de estabilidade.

Esse comportamento levanta a discussão sobre o equilíbrio entre representatividade física e desempenho estatístico: modelos mais gerais tendem a apresentar menor ajuste, enquanto modelos mais restritivos exibem maior precisão, porém aplicáveis a um intervalo reduzido de condições que explicam um mesmo contexto geofísico.

7. CONCLUSÕES

Este trabalho avaliou o desempenho de modelos de regressão linear múltipla aplicados à estimativa do Fator de Segurança de taludes, considerando diferentes níveis de tratamento de dados. A comparação entre os três cenários demonstrou que a qualidade e o pré-processamento do banco de dados exercem influência direta no ajuste estatístico, no poder explicativo e na confiabilidade dos modelos.

O modelo construído a partir dos dados brutos apresentou baixo desempenho, refletindo a elevada variabilidade natural dos parâmetros geotécnicos e a influência de valores extremos, que dificultaram a identificação de relações lineares mais consistentes. A remoção de *outliers* multivariados por meio da distância de Mahalanobis resultou em melhora moderada do segundo ajuste, evidenciando que os *outliers* exerciam impacto significativo sobre os coeficientes de regressão estimados.

O melhor desempenho foi obtido no cenário em que o banco de dados foi previamente tratado baseado em critérios de valores típicos para as variáveis independentes. A exclusão de observações inconsistentes e de condições extremas promoveu a homogeneização do conjunto de dados, e melhor ajuste do modelo. Os valores obtidos para o R^2 ajustado, juntamente com os resultados da validação cruzada, indicaram boa capacidade de generalização dentro do domínio analisado.

Entretanto, é válido ressaltar que a maior confiabilidade estatística do último cenário está associada a restrição do domínio de aplicação do modelo. Apesar da redução da variabilidade global dos dados, foi preservada uma variabilidade representativa do comportamento típico de solos residuais, tornando o modelo adequado para taludes artificiais e encostas em contextos urbanos, rodoviários e ferroviários, que se enquadram nas restrições adotadas.

Dessa forma, este modelo se mostra mais adequado para análises preliminares e estimativas médias, devendo ser utilizado de forma complementar aos métodos determinísticos tradicionais em avaliações de estabilidade de taludes nestes tipos de solo.

Em trabalhos futuros, sugere-se a ampliação do banco de dados, a avaliação de modelos não lineares e a aplicação da metodologia em casos reais para validação externa, buscando ampliar o domínio de validade e a aplicabilidade prática dos modelos propostos.

REFERÊNCIAS BIBLIOGRÁFICAS

ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6484:** Solo – Sondagens de simples reconhecimento com SPT – Método de ensaio. Rio de Janeiro, 2020.

AHMAD, Feezan et al. **Prediction of slope stability using Tree Augmented Naive-Bayes classifier:** Modeling and performance evaluation. *Math. Biosci. Eng.*, v. 19, n. 5, p. 4526-4546, 2022.

CINTRA, José Carlos A.; AOKI, Nelson; ALBIERO, José H. **Fundações Diretas: Projeto Geotécnico.** Porto Alegre: Oficina de Texto, 2025. E-book. p.17. ISBN 9788579750878. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788579750878/>. Acesso em: 11 jan. 2026.

DAS, Braja M.; SOBHAN, Khaled. **Fundamentos de engenharia geotécnica.** 4. ed. Porto Alegre: +A Educação - Cengage Learning Brasil, 2019. E-book. p.397. ISBN 9788522128280. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788522128280/>. Acesso em: 09 jan. 2026.

DUTT-ROSS, Steven. **Manual de análise de dados.** 23 abr. 2020. Atualizado em: 21 jul. 2020. Disponível em: < <https://livro.metodosquantitativos.com/docs/index.html> >. Acesso em: 14 jan. 2026.

FERNANDES, Manuel de Matos. **Mecânica dos solos:** introdução à engenharia geotécnica. 1. ed. São Paulo: Grupo A, 2014. E-book. Disponível em: <https://plataforma.bvirtual.com.br>. Acesso em: 11 jan 2026.

FIORI, Alberto Pio. **Fundamentos de mecânica dos solos e das rochas:** aspectos geotécnicos. 3. ed. São Paulo: Grupo A, 2015. E-book. Disponível em: <https://plataforma.bvirtual.com.br>. Acesso em: 09 jan 2026.

GERSCOVICH, D. M. S. **Estabilidade de taludes.** 2. ed. São Paulo: Grupo A, 2016. E-book. Disponível em: <https://plataforma.bvirtual.com.br>. Acesso em: 09 jan 2026.

HOQUE, Md. J.; BAYEZID, Md.; SHARAN, A. R.; UL KABIR, M.; TAREQUE, T. **Prediction of strength properties of soft soil considering simple soil parameters.** Open Journal of Civil Engineering, v. 13, p. 479–496, 2023. DOI: 10.4236/ojce.2023.133035.

JR., Joseph F H.; BLACK, William C.; BABIN, Barry J.; et al. **Análise multivariada de dados.** 6. ed. Porto Alegre: Bookman, 2009. E-book. p.169. ISBN 9788577805341. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788577805341/>. Acesso em: 12 jan. 2026.

LUGU, Benjamin Kweku. **Exploratory Factor Analysis in R.** Bookdown, 2023. Disponível em: https://bookdown.org/luguben/EFA_in_R/index.html. Acesso em: 09 fev. 2026.

MAIA, Alexandre Gori. **Régressão linear múltipla: econometria.** Notas de aula. Universidade Estadual de Campinas (UNICAMP), [s.d.].

MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística Aplicada e Probabilidade para Engenheiros.** 7. ed. Rio de Janeiro: LTC, 2021. E-book. p.234. ISBN 9788521637448. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788521637448/>. Acesso em: 25 jan. 2026.

MURRIETA, Pedro. **Mecânica dos Solos.** Rio de Janeiro: GEN LTC, 2018. E-book. p.327. ISBN 9788595156074. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788595156074/>. Acesso em: 09 jan. 2026.

PINTO, Carlos de S. **Curso Básico de Mecânica dos Solos em 16 Aulas.** 3. ed. Porto Alegre: Oficina de Texto, 2025. E-book. p.4. ISBN 9788579751165. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788579751165/>. Acesso em: 09 jan. 2026.

POSIT. **RStudio Desktop** – download. Disponível em: <<https://posit.co/download/rstudio-desktop/>>. Acesso em: 14 jan. 2026.

PRIEBE, Edson Ricardo; SCHEIBNER, Djhony Charles; ALMEIDA, Maycon André de. **Análise numérica de contenção em parede diafragma para quatro subsolos em Cascavel/PR.** In: SIMPÓSIO DE PRÁTICAS DE ENGENHARIA GEOTÉCNICA DA REGIÃO SUL (GEOSUL), 12., 2019, Joinville. Anais [...]. Joinville: ABMS, 2019.

SANTOS, Allan Erlikman Medeiros. **Predição da condição de estabilidade de taludes de mina por meio de estatística multivariada.** Dissertação (Mestrado em Engenharia Mineral) – Universidade Federal de Ouro Preto, Ouro Preto, 2016.

SERVIÇO GEOLOGICO DO BRASIL — CPRM. **Diagnóstico da população em áreas de risco geológico: Ouro Preto, MG.** Brasília: CPRM, 2021. 64 p. Disponível em:https://rigeo.sgb.gov.br/bitstream/doc/22239/3/diagnostico_risco_ouro_preto_mg.pdf. Acesso em: 26 jan. 2026.

WYLLIE, Duncan C.; MAH, Chris W. **Rock slope engineering: civil and mining.** 4. ed. London; New York: Spon Press, 2004.

APÊNDICES

APÊNDICE A – ALGORITMO ANÁLISE 1

```

##ANÁLISE 1

##LEITURA DOS DADOS
dados <- read.table("database.txt",header=TRUE, row.names = 1, dec = '.')
dados

## dimensão dos dados
dim(dados)

## estrutura do conjunto de dados
str(dados)

## Vetor de médias
colMeans(dados)

## Matriz de Covariâncias
S = var(dados)
S

## Matriz de Correlação dos Dados
R= cor(dados)
R

##Resumo Estatísticos
summary(dados)

##BOXPLOT
x11()
boxplot(dados, main = "Boxplot das variáveis")
boxplot(dados$γ, xlab = "Peso Específico", ylab = "kNm-3", main = "Boxplot da variável Peso Específico")
boxplot(dados$c, xlab = "Coesão", ylab = "kPa", main = "Boxplot da variável Coesão")
boxplot(dados$φ, xlab = "Ângulo de atrito", ylab = "graus", main = "Boxplot da variável Ângulo de Atrito")

```

```

boxplot(dados$β, xlab = "Ângulo de Inclinação", ylab = "graus", main =
"Boxplot da variável Ângulo de Inclinação")

boxplot(dados$H, xlab = "Altura", ylab = "metros", main = "Boxplot da
variável Altura")

boxplot(dados$ru, xlab = "Razão de Poropressão", main = "Boxplot da
variável Razão de Poropressão")

boxplot(dados$FS, xlab = "Fator de Segurança", main = "Boxplot da variável
Fator de Segurança")

## HISTOGRAMA

hist(dados$γ, xlab = "Peso Específico (kN/m³)", ylab = "Frequência", main =
"Histograma da variável Peso Específico")

hist(dados$c, xlab = "Coesão (kPa)", ylab = "Frequência", main =
"Histograma da variável Coesão")

hist(dados$φ, xlab = "Ângulo de atrito (graus)", ylab = "Frequência", main =
"Histograma da variável Ângulo de atrito")

hist(dados$β, xlab = "Ângulo de Inclinação (graus)", ylab = "Frequência", main =
"Histograma da variável Ângulo de Inclinação")

hist(dados$H, xlab = "Altura (m)", ylab = "Frequência", main = "Histograma
da variável Altura")

hist(dados$ru, xlab = "Razão de Poropressão", ylab = "Frequência", main =
"Histograma da variável Razão de Poropressão")

hist(dados$FS, xlab = "Fator de Segurança", ylab = "Frequência", main =
"Histograma da variável Fator de Segurança")

## scatterplot matriz

pairs(dados)

## scatterplot matriz com correlações

## função para personalização do painel

painel.pearson <- function(x, y, ...) {
  horizontal <- (par("usr")[1] + par("usr")[2]) / 2;
  vertical <- (par("usr")[3] + par("usr")[4]) / 2;
  text(horizontal, vertical, format(abs(cor(x,y))), digits=2), cex = 1.2,
  font = 1)
}

pairs(dados, main = "Matriz de Dispersão - Análise 1", pch = 21,
      upper.panel = painel.pearson)

## Teste de esfericidade de Bartlett

library(psych)

```

```
n=dim(dados) [1]
cortest.bartlett(R,n)

## Teste de normalidade multivariada
library(MVN)
mvn(dados)

##REGRESSÃO LINEAR MULTIPLA
modelo <- lm(dados$FS ~., data = dados)
summary(modelo)

## Homocedasticidade dos resíduos
plot(rstudent(modelo) ~ fitted(modelo), pch = 19)
abline(h = 0, lty = 2, col = "red")

## Normalidade dos Resíduos
hist(x = modelo$residuals, col = 'gray', xlab = 'Resíduos', ylab =
'Densidade de Probabilidade',
probability = TRUE)
lines(density(modelo$residuals))
```

APÊNDICE B – ALGORITMO ANÁLISE 2

```
##ANÁLISE 2

##LEITURA DOS DADOS

dados <- read.table("database.txt",header=TRUE, row.names = 1,
dec = '.')

dados

## dimensao dos dados

dim(dados)

## estrutura do conjunto de dados

str(dados)

## Vetor de médias

colMeans(dados)

## Matriz de Covariâncias

S = var(dados)

S

## Matriz de Correlação dos Dados

R= cor(dados)

R

##ESTATISTICA DESCRIPTIVA

summary(dados)

##BOXPLOT

x11()
```

```
boxplot(dados, main = "Boxplot das variáveis")
boxplot(dados$γ, xlab = "Peso Específico", ylab = "kNm-3",
main = "Boxplot da variável Peso Específico")
boxplot(dados$c, xlab = "Coesão", ylab = "kPa", main =
"Boxplot da variável Coesão")
boxplot(dados$φ, xlab = "Ângulo de atrito", ylab = "graus",
main = "Boxplot da variável Ângulo de Atrito")
boxplot(dados$β, xlab = "Ângulo de Inclinação", ylab =
"graus", main = "Boxplot da variável Ângulo de Inclinação")
boxplot(dados$H, xlab = "Altura", ylab = "metros", main =
"Boxplot da variável Altura")
boxplot(dados$ru, xlab = "Razão de Poropressão", main =
"Boxplot da variável Razão de Poropressão")
boxplot(dados$FS, xlab = "Fator de Segurança", ylab = "kPa",
main = "Boxplot da variável Fator de Segurança")

## HISTOGRAMA

hist(dados$γ, xlab = "Peso Específico", ylab = "Frequência",
main = "Histograma da variável Peso Específico")
hist(dados$c, xlab = "Coesão", ylab = "Frequência", main =
"Histograma da variável Coesão")
hist(dados$φ, xlab = "Ângulo de atrito", ylab = "Frequência",
main = "Histograma da variável Ângulo de atrito")
hist(dados$β, xlab = "Ângulo de Inclinação", ylab =
"Frequência", main = "Histograma da variável Ângulo de
Inclinação")
hist(dados$H, xlab = "Altura", ylab = "Frequência", main =
"Histograma da variável Altura")
hist(dados$ru, xlab = "Razão de Poropressão", ylab =
"Frequência", main = "Histograma da variável Razão de
Poropressão")
hist(dados$FS, xlab = "Fator de Segurança", ylab =
"Frequência", main = "Histograma da variável Fator de
Segurança")
```

```
## scatterplot matriz
pairs(dados) #verificaçāo da dispersāo de todos os dados

## scatterplot matrix com correlacōes
## funcao para personalizacāo do painel
painel.pearson <- function(x, y, ...) {
  horizontal <- (par("usr")[1] + par("usr")[2]) / 2;
  vertical <- (par("usr")[3] + par("usr")[4]) / 2;
  text(horizontal, vertical, format(abs(cor(x,y))), digits=2),
  cex = 1.2,
  font = 1)
}
pairs(dados, main = "Conjunto de Dados", pch = 21,
      upper.panel = painel.pearson)

## Teste de esfericidade de Bartlett
library(psych)
n=dim(dados)[1]
cortest.bartlett(R,n)

## Teste de normalidade multivariada
library(MVN)
mvn(dados)

##Detecçāo de outliers multivariados (distancia de
mahalanobis)
maha2 <- mahalanobis(dados, center = colMeans(dados), cov =
cov(dados))
maha2
```

```
# calcular o c^2, n graus de liberdade
quant2 <- qchisq(0.95, dim(dados) [2])
quant2
x11()
plot(1:length(maha2), maha2, xlab = "Elementos amostrais",
      ylab= "Distância de Mahalanobis")
abline(h=quant2, lty=2, col="red")
out2 <- maha2[maha2>quant2]
out2
out2 = as.data.frame(out2)
out2
dim(out2)
text(row.names(out2),out2$out2, row.names(out2), pos = 2 )

## Remoção dos outliers do banco de dados
ifelse(maha2>quant2, FALSE, TRUE)
row_to_keep = ifelse(maha2>quant2, FALSE, TRUE)
dados_s_out = dados[row_to_keep,]
dados_s_out

#ESTATÍSTICA DESCRIPTIVA PÓS REMOÇÃO OUTLIERS
summary(dados_s_out)
boxplot(dados_s_out, main = "Boxplot das variáveis")

#REGRESSÃO LINEAR MULTIPLA
modelo <- lm(dados_s_out$FS ~., data = dados_s_out)
summary(modelo)
```

```
## Homocedasticidade dos resíduos
x11()
plot(rstudent(modelo) ~ fitted(modelo), pch = 19)
abline(h = 0, lty = 2, col = "red")

## Normalidade dos Resíduos
hist(x = modelo$residuals, col = 'gray', xlab = 'Resíduos',
ylab = 'Densidade de Probabilidade',
probability = TRUE)
lines(density(modelo$residuals), col="red")
```

APÊNDICE C – ALGORITMO ANÁLISE 3

```
##ANÁLISE 3

##LEITURA DOS DADOS

dados <- read.table("database_tratado.txt", header=TRUE, row.names = 1, dec
= '.')

dados

## dimensao dos dados

dim(dados) ## n = 40, p = 7

## estrutura do conjunto de dados

str(dados)

## Vetor de médias

colMeans(dados)

## Matriz de Covariâncias

S = var(dados)

S

## Matriz de Correlação dos Dados

R= cor(dados)

R

##ESTATISTICA DESCRIPTIVA

summary(dados)

##BOXPLOT

x11()

boxplot(dados, main = "Boxplot das variáveis")

boxplot(dados$y, xlab = "Peso Específico", ylab = "kNm-3", main = "Boxplot
da variável Peso Específico")
```

```

boxplot(dados$c, xlab = "Coesão", ylab = "kPa", main = "Boxplot da variável
Coesão")

boxplot(dados$φ, xlab = "Ângulo de atrito", ylab = "graus", main = "Boxplot
da variável Ângulo de Atrito")

boxplot(dados$β, xlab = "Ângulo de Inclinação", ylab = "graus", main =
"Boxplot da variável Ângulo de Inclinação")

boxplot(dados$H, xlab = "Altura", ylab = "metros", main = "Boxplot da
variável Altura")

boxplot(dados$ru, xlab = "Razão de Poropressão", main = "Boxplot da
variável Razão de Poropressão")

boxplot(dados$FS, xlab = "Fator de Segurança", ylab = "kPa", main =
"Boxplot da variável Fator de Segurança")

## HISTOGRAMA

hist(dados$γ, xlab = "Peso Específico (kN/m³)", ylab = "Frequência", main =
"Histograma da variável Peso Específico")

hist(dados$c, xlab = "Coesão (kPa)", ylab = "Frequência", main =
"Histograma da variável Coesão")

hist(dados$φ, xlab = "Ângulo de atrito (graus)", ylab = "Frequência", main =
"Histograma da variável Ângulo de atrito")

hist(dados$β, xlab = "Ângulo de Inclinação (graus)", ylab = "Frequência",
main = "Histograma da variável Ângulo de Inclinação")

hist(dados$H, xlab = "Altura (m)", ylab = "Frequência", main = "Histograma
da variável Altura")

hist(dados$ru, xlab = "Razão de Poropressão", ylab = "Frequência", main =
"Histograma da variável Razão de Poropressão")

hist(dados$FS, xlab = "Fator de Segurança", ylab = "Frequência", main =
"Histograma da variável Fator de Segurança")

## scatterplot matrix

x11()
pairs(dados)

## scatterplot matrix com correlações

## função para personalização do painel
painele.pearson <- function(x, y, ...) {
  horizontal <- (par("usr")[1] + par("usr")[2]) / 2;
  vertical <- (par("usr")[3] + par("usr")[4]) / 2;
}

```

```

text(horizontal, vertical, format(abs(cor(x,y))), digits=2), cex = 1.2,
font = 1)
}

pairs(dados, main = "Conjunto de Dados", pch = 21,
upper.panel = painel.pearson)

## Teste de espericidade de Bartlett
library(psych)
n=dim(dados) [1]
cortest.bartlett(R,n)

## Teste de normalidade multivariada
library(MVN)
mvn(dados)

####Detecção de outliers multivariados (distancia de mahalanobis)
maha2 <- mahalanobis(dados, center = colMeans(dados), cov = cov(dados))
maha2

# para calcular o c^2, n graus de liberdade
quant2 <- qchisq(0.95, dim(dados) [2])
quant2
plot(1:length(maha2), maha2, xlab = "Elementos amostrais",
ylab= "Distância de Mahalanobis")
abline(h=quant2, lty=2, col="red")
out2 <- maha2[maha2>quant2]
out2
out2 = as.data.frame(out2)
out2
dim(out2)

text(row.names(out2),out2$out2,row.names(out2), pos = 2 )

```

```

## Removendo os outliers do banco de dados
ifelse(maha2>quant2, FALSE, TRUE)
row_to_keep = ifelse(maha2>quant2, FALSE, TRUE)
dados_s_out = dados[row_to_keep,]
dados_s_out
dim(dados_s_out)

#REGRESSÃO LINEAR MULTIPLA
modelo <- lm(dados_s_out$FS ~., data = dados_s_out)
summary(modelo)

## Homocedasticidade dos resíduos
plot(rstudent(modelo) ~ fitted(modelo), pch = 19)
abline(h = 0, lty = 2, col = "red")

## Normalidade dos Resíduos
hist(x = modelo$residuals, col = 'gray', xlab = 'Resíduos', ylab =
'Densidade de Probabilidade',
probability = TRUE)
lines(density(modelo$residuals), col="red")

##Validação Cruzada
kfold_cv_lm <- function(dados, k){

  set.seed(123)  # reproducibilidade
  n <- nrow(dados)
  folds <- sample(rep(1:k, length.out = n))

  rmse <- numeric(k)
  r2    <- numeric(k)
  for(i in 1:k){

    treino <- dados[folds != i, ]
    teste  <- dados[folds == i, ]
  }
}

```

```

modelo_cv <- lm(FS ~ ., data = treino)

pred <- predict(modelo_cv, newdata = teste)

rmse[i] <- sqrt(mean((teste$FS - pred)^2))

r2[i]    <- cor(teste$FS, pred)^2

}

return(list(
  RMSE_medio = mean(rmse),
  RMSE_sd     = sd(rmse),
  R2_medio   = mean(r2),
  R2_sd      = sd(r2)
))

}

## K=5

cv_5 <- kfold_cv_lm(dados_s_out, k = 5)
cv_5

## K=10

cv_10 <- kfold_cv_lm(dados_s_out, k = 10)
cv_10

##TABELA COMPARATIVA CV

resultado_cv <- data.frame(
  K = c(5, 10),
  RMSE_medio = c(cv_5$RMSE_medio, cv_10$RMSE_medio),
  RMSE_sd    = c(cv_5$RMSE_sd, cv_10$RMSE_sd),
  R2_medio   = c(cv_5$R2_medio, cv_10$R2_medio),
  R2_sd      = c(cv_5$R2_sd, cv_10$R2_sd)
)

resultado_cv

```

ANEXOS

ANEXO A – BANCO DE DADOS

Tabela A1 – Banco de dados com os parâmetros geomecânicos utilizados.

No.	γ/kNm^{-3}	c/kPa	$\phi /{}^\circ$	$\beta /{}^\circ$	H/m	r_u	FoS	SS
1	14	11.97	26	30	88	0.45	0.625	0
2	27	37.5	35	37.8	320	0.25	1.24	1
3	12	0	30	35	4	0	1.46	1
4	22.4	10	35	45	10	0.4	0.9	0
5	21	35	28	40	12	0.5	1.43	1
6	20	10.1	29	34	6	0.3	1.34	1
7	27	40	35	47.1	292	0.25	1.15	0
8	28.4	29.4	35	35	100	0	1.78	1
9	27.3	31.5	29.7	41	135	0.25	1.245	1
10	22	20	22	20	180	0.1	0.99	0
11	22.4	10	35	30	10	0	2	1
12	27.3	10	39	41	511	0.25	1.434	1
13	19	30	35	35	11	0.2	2	1
14	27.3	10	39	40	470	0.25	1.418	1
15	14	12	26	30	88	0	1.02	0
16	19.1	10.1	10	25	50	0.4	0.65	0
17	18.7	26.4	15	35	8.2	0	1.11	0
18	20	0	36	45	50	0.25	0.79	0
19	22	20	22	20	180	0	1.12	0
20	19.6	12	20	22	12.2	0.405	1.35	0
21	16	70	20	40	115	0	1.11	0
22	19	11.7	28	35	21	0.11	1.09	0
23	21	45	25	49	12	0.3	1.53	1
24	20	20	36	45	50	0.5	0.83	0
25	18.8	30	20	30	50	0.1	1.46	1
26	14.8	0	17	20	50	0	1.13	0
27	27	35	35	42	359	0.25	1.27	1
28	20	0	24.5	20	8	0.35	1.37	1
29	18	24	30.2	45	20	0.12	1.12	0
30	25	46	36	44.5	299	0.25	1.55	1
31	27	32	33	42.4	289	0.25	1.3	1
32	22	0	36	45	50	0	0.89	0
33	18.8	20	10	25	50	0.3	0.97	0
34	18.8	25.1	20	30	50	0.2	1.21	0
35	27.3	10	39	40	480	0.25	1.45	1
36	27.3	16.8	28	50	90.5	0.25	1.252	1
37	20	40.1	30	30	15	0.3	1.84	1
38	18.8	14.4	25	20	30.6	0	1.88	1
39	21.5	6.9	30	31	76.8	0.38	1.01	0
40	14	11.97	26	30	88	0	1.02	0
41	26	150	45	50	200	0	1.2	1
42	25	46	35	46	432	0.25	1.23	1

43	18.5	12	0	30	6	0	0.78	0
44	18	45	25	25	14	0.3	2.09	1
45	22.4	100	45	45	15	0.25	1.8	1
46	20.6	16.2	26.5	30	40	0	1.25	0
47	25	46	35	50	284	0.25	1.34	1
48	18.8	20	20	30	50	0.3	1	0
49	21	20	40	40	12	0	1.84	1
50	18.8	25.1	10	25	50	0.2	1.18	0
51	23.47	0	32	37	214	0	1.08	0
52	21.43	0	20	20	61	0.5	1.03	0
53	18.5	25	0	30	6	0	1.09	0
54	31.3	68	37	49	200.5	0.25	1.2	0
55	28.4	39.2	38	35	100	0	1.99	1
56	18.8	14.4	25	20	30.6	0.45	1.11	0
57	27.3	14	31	41	110	0.25	1.249	1
58	31.3	68	37	46	366	0.25	1.2	0
59	20	40.1	40	40	10	0.2	2.31	1
60	21.8	8.6	32	28	12.8	0.49	1.03	0
61	18.8	30	10	25	50	0.1	1.4	1
62	18.84	0	20	20	7.62	0.45	1.05	0
63	18.8	10.4	21.3	34	37	0.3	1.29	0
64	20.4	24.9	13	22	10.6	0.35	1.4	1
65	27	32	33	42.6	301	0.25	1.16	0
66	22	0	40	33	8	0.35	1.45	1
67	21.4	10	30.34	30	20	0	1.7	1
68	20	0	36	45	50	0.5	0.67	0
69	16.5	11.6	0	30	3.6	0	1	0
70	18.8	57.5	20	20	30.6	0	2.04	1
71	12	0	30	45	8	0	0.8	0
72	18	5	30	20	8	0.3	2.05	1
73	18.84	14.36	25	20	30.5	0.45	1.11	0
74	19.1	10.1	20	30	50	0.4	0.65	0
75	25	46	35	47	443	0.25	1.28	1
76	18.8	24.8	21.3	29.2	37	0.5	1.07	0
77	22	20	36	45	50	0	1.02	0
78	25	120	45	53	120	0	1.3	1
79	23	0	20	20	100	0.3	1.2	0
80	20.4	33.5	11	16	45.8	0.2	1.28	0
81	25	46	35	44	435	0.25	1.37	1
82	18.8	15.3	30	25	10.6	0.38	1.63	1
83	21	30	35	40	12	0.4	1.49	1
84	24	0	40	33	8	0.3	1.58	1
85	14	12	26	30	88	0.45	0.63	0
86	20	20	36	45	50	0.25	0.96	0
87	27.3	26	31	50	92	0.25	1.246	1

Fonte: Ahmad et al (2022).