

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

Instituto de Ciências Exatas

Curso de Especialização em Estatística

Amanda Queiroz de Paula

**Avaliação do modelo recorrente de regressão de aceleração de pico do terreno  
com base em dados de uma mineradora do interior de São Paulo, Brasil**

Belo Horizonte

2022

Amanda Queiroz de Paula

**Avaliação do modelo recorrente de regressão de aceleração de pico do terreno  
com base em dados de uma mineradora do interior de São Paulo, Brasil**

**Versão final**

Monografia de especialização apresentada  
ao Instituto de Ciências Exatas da  
Universidade Federal de Minas Gerais,  
como requisito parcial para a obtenção do  
título de Especialista em Estatística.

Área de concentração: Estatística

Orientador: Prof. Dr. Roberto Quinino

Belo Horizonte

2022

Paula, Amanda Queiroz de.

P324a Avaliação do modelo recorrente de regressão de aceleração de pico do terreno com base em dados de uma mineradora do interior de São Paulo, Brasil [manuscrito] / Amanda Queiroz de Paula – 2022.  
47 f. il.

Orientador: Roberto da Costa Quinino.

Monografia(especialização) - Universidade Federal de Minas Gerais Instituto de Ciências Exatas, Departamento de Estatística.

Referências: f. 46-47.

1. Estatística. 2. Análise de regressão. 3. Sismologia – Métodos estatísticos. 4. Modelo de movimentação do terreno – Teoria da previsão. I. Quinino, Roberto da Costa. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2(043)



**Universidade Federal de Minas Gerais**  
**Instituto de Ciências Exatas**  
**Departamento de Estatística**  
**Programa de Pós-Graduação / Especialização**  
Av. Pres. Antônio Carlos, 6627 - Pampulha  
31270-901 – Belo Horizonte – MG

E-mail: [pgest@ufmg.br](mailto:pgest@ufmg.br)  
Tel: 3409-5923 – FAX: 3409-5924

## **ATA DO 250ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE AMANDA QUEIROZ DE PAULA.**

Aos quatorze dias do mês de novembro de 2022, às 17:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso da aluna **Amanda Queiroz de Paula**, intitulado: “*Avaliação do modelo recorrente de regressão de aceleração de pico do terreno com base em dados de uma mineradora do interior de São Paulo, Brasil*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Roberto da Costa Quinino – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: a candidata foi considerada Aprovada condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 14 de novembro de 2022.

  
Prof. Roberto da Costa Quinino (Orientador)  
**Departamento de Estatística / ICEX / UFMG**

Documento assinado digitalmente  
gov.br ROBERTO DA COSTA QUININO  
Data: 14/11/2022 18:53:27-0300  
Verifique em <https://verificador.iti.br>

ANDERSON LAECIO GALINDO  
TRINDADE:00954199693

Assinado de forma digital por  
ANDERSON LAECIO GALINDO  
TRINDADE:00954199693  
Dados: 2022.11.17 09:40:31 -03'00'

Prof. Anderson Laécio Galindo Trindade  
**DEP / UFMG**



**Universidade Federal de Minas Gerais**  
**Instituto de Ciências Exatas**  
**Departamento de Estatística**  
**Programa de Pós-Graduação / Especialização**  
Av. Pres. Antônio Carlos, 6627 - Pampulha  
31270-901 – Belo Horizonte – MG

E-mail: [pgest@ufmg.br](mailto:pgest@ufmg.br)  
Tel: 3409-5923 – FAX: 3409-5924

## **DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.**

Declaro para os devidos fins que Amanda Queiroz de Paula, número de registro 2018696534, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística e que me entregou a versão final corrigida. O trabalho foi apresentado no dia 14 de novembro de 2022 com o título “*Avaliação do modelo recorrente de regressão de aceleração de pico do terreno com base em dados de uma mineradora do interior de São Paulo, Brasil*”.

Belo Horizonte, 16 de novembro de 2022

Roberto da Costa  
Quinino:80871291720  
Assinado de forma digital por Roberto da Costa Quinino:80871291720  
Dados: 2022.11.16 19:07:11 -03'00'

*Prof. Roberto da Costa Quinino*  
Coordenador da Comissão do Curso de Especialização em Estatística  
Prof. Roberto da Costa Quinino  
Coordenador do curso de Especialização em Estatística  
Departamento de Estatística / UFMG

## RESUMO

A sismicidade é intrínseca à existência do nosso planeta, de forma que o entendimento da sismologia é imprescindível para a humanidade. Modelos de movimentação de terra (GMM) buscam estimar a vibração em determinado ponto a partir de dados da fonte sísmica e sua distância. Dentre as muitas aplicações, pode-se apontar, por exemplo, o projeto de estruturas resistentes a terremotos. Na literatura sismológica, são recorrentes GMM que considerem um modelo de regressão múltipla linear a partir do logaritmo da aceleração de pico do terreno (PGA) sendo estimada a partir dos logaritmos de potência sísmica e distância da fonte. O objetivo desse estudo foi avaliar aspectos estatísticos sobre essa abordagem, perspectiva comumente negligenciada na literatura específica. Foram elaborados cinco (05) modelos: dois de regressão linear simples (PGA em função de potência e de distância separadamente) e três de regressão linear múltipla (com dados com transformação logarítmica e sem transformação). Os resultados mostraram que, para os dados considerados (intervalo de um ano de um sistema de monitoramento microsísmico de cava de São Paulo - Brasil), as premissas de normalidade e homoscedasticidade dos resíduos não foram respeitadas em nenhum modelo. Os coeficientes de determinação preditivos não ultrapassaram 16%, valor baixo, mas possivelmente explicado por: (i) falta de distinção da fonte sísmica na base de dados, (ii) heterogeneidade da geologia ao redor da cava, (iii) falhas no sistema de monitoramento microsísmico e (iv) limitações na abordagem estatísticas. Portanto, o modelo amplamente utilizado na Sismologia, para o banco de dados empregado, não se mostrou estatisticamente válido. Sugere-se a abordagem não-linear para se tentar solucionar a questão.

**Palavras-chave:** Modelo de movimentação do terreno. Equações de predição de movimentação de terreno. Monitoramento microsísmico. Aceleração de pico do terreno. Terremotos.

## ABSTRACT

Seismicity is intrinsic to the existence of our planet, so the understanding of seismology is essential for humanity. Ground motion models (GMM) seek to estimate the vibration at a given point from seismic source data and its distance. Among the many applications, one example is the design of earthquake reinforced structures. In the seismological literature, GMMs that consider a linear multiple regression model from the logarithm of peak ground acceleration (PGA) are recurrent, being estimated from the logarithms of seismic power and distance from the source. The objective of this study was to evaluate statistical aspects of this approach, a perspective commonly neglected in the specific literature. Five (05) models were developed: two of simple linear regression (PGA as a function of power and distance separately) and three of multiple linear regression (with data with and without logarithmic transformation). The results suggested that, for the considered data (one year period of a microseismic monitoring system of an open pit in São Paulo - Brazil), the assumptions of normality and homoscedasticity of the residues were not respected by any model. The adjusted determination coefficients did not exceed 16%, a low value, but possibly explained by: (i) lack of distinction of seismic sources in the database, (ii) heterogeneity of the geology around the pit, (iii) failures in the microseismic monitoring system and (iv) limitations in the statistical approach. Therefore, such model widely used in Seismology, for the employed database, was not statistically valid. A non-linear approach is suggested as an alternative to solve the issue.

**Key words:** Ground motion models. Ground motion predictive equations. Microseismic monitoring. Peak ground acceleration. Earthquakes.

## LISTA DE ILUSTRAÇÕES

Figura 2.1 – Esquema de diferentes tipos de fontes sísmicas.....	13
Figura 2.2 – Esquema de sismograma típico com apontamento do PGA. ....	14
Figura 2.3 – Ilustração da distância epicentral, adotada no presente trabalho.....	15
Figura 3.1 – Localização dos geofones ao redor da cava monitorada. ....	17
Figura 4.1 – Resumo estatístico da variável independente “PGA”.....	19
Figura 4.2 – Histograma da variável dependente “PGA” com truncamento superior para melhor visualização de maior parte dos indivíduos. ....	20
Figura 4.3 – Resumo estatístico da variável independente “distância”. ....	21
Figura 4.4 – Resumo estatístico da variável independente “potência”.....	22
Figura 4.5 – Histograma da variável independente “potência sísmica” com truncamento superior para melhor visualização de maior parte dos indivíduos. ....	23
Figura 4.6 – Matriz de dispersão pareada entre as três variáveis.....	24
Figura 4.7 – Dados da matriz de dispersão pareada entre as três variáveis. ....	24
Figura 4.8 – Modelo de regressão linear simples para PGA predito por distância.....	26
Figura 4.9 – Regressão simples entre PGA e distância para os dados considerados. ....	27
Figura 4.10 – Análise dos resíduos padronizados para o modelo de regressão linear simples PGA x distância.....	27
Figura 4.11 – Modelo de regressão linear simples para PGA predito por potência...28	28
Figura 4.12 – Regressão simples entre PGA e potência sísmica para os dados considerados. ....	29
Figura 4.13 – Análise dos resíduos padronizados para o modelo de regressão linear simples PGA x potência sísmica.....	29
Figura 4.14 – Modelo de regressão linear simples para PGA predito tanto por distância, quanto por potência.....	31
Figura 4.15 – Análise dos resíduos padronizados para o modelo de regressão linear simples para PGA a partir de potência sísmica e distância.....	32
Figura 4.16 – Histograma da variável dependente “PGA” face a transformação logarítmica. ....	33
Figura 4.17 – Histograma da variável dependente “PGA” face a transformação Box-Cox.....	34

Figura 4.18 – Histograma da variável dependente “PGA” face a transformação de Johnson.....	34
Figura 4.19 – Resumo estatístico da variável independente “logPGA” .....	35
Figura 4.20 – Resumo estatístico da variável independente “log da distância” .....	36
Figura 4.21 – Resumo estatístico da variável independente “log da potência” .....	36
Figura 4.22 – Matriz de correlação, entre as três variáveis transformadas por algoritmo.....	37
Figura 4.23 – Modelo de regressão linear simples para log do PGA predito tanto por log da distância, quanto por log da potência.....	38
Figura 4.24 – Análise dos resíduos padronizados para o modelo de regressão linear simples para log de PGA a partir de log da potência sísmica e log da distância.....	39
Figura 4.25 – Matriz de correlação, entre as log(PGA), log(d) e potência.....	40
Figura 4.26 – Modelo de regressão linear simples para log do PGA predito por log da distância e potência.....	41
Figura 4.27 – Análise dos resíduos padronizados para o modelo de regressão linear simples para log de PGA a partir de log da potência sísmica e log da distância.....	42

## LISTA DE TABELA

Tabela 1 – Compilação dos coeficientes de determinação preditivos dos modelos construídos .....	44
--	----

# SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>12</b>
<b>2. CONTEXTUALIZAÇÃO TEÓRICA.....</b>	<b>13</b>
2.1. ACELERAÇÃO DE PICO DE TERRENO [M/S <sup>2</sup> ] .....	14
2.2. POTÊNCIA SÍSMICA [M <sup>3</sup> ].....	14
2.3. DISTÂNCIA [M] .....	15
2.4. MODELOS DE MOVIMENTAÇÃO DE TERRENO .....	15
<b>3. MÉTODOS.....</b>	<b>17</b>
3.1. COLETA E SELEÇÃO DE DADOS .....	17
3.2. BANCO DE DADOS EMPREGADO.....	18
3.3. SELEÇÃO DE VARIÁVEIS .....	18
<b>4. RESULTADOS .....</b>	<b>19</b>
4.1. ANÁLISE EXPLORATÓRIA .....	19
4.1.1. <i>Aceleração de pico do terreno</i> .....	19
4.1.2. <i>Distância</i> .....	21
4.1.3. <i>Potência sísmica</i> .....	22
4.1.4. <i>Correlação entre as variáveis</i> .....	23
4.2. REGRESSÕES LINEARES SIMPLES.....	25
4.2.1. <i>PGA versus distância</i> .....	25
4.2.2. <i>PGA versus potência sísmica</i> .....	28
4.3. REGRESSÃO LINEAR MÚLTIPLA .....	30
4.3.1. <i>Avaliação de multicolinearidade</i> .....	30
4.3.2. <i>Resultado da regressão linear múltipla</i> .....	30
4.4. TRANSFORMAÇÃO DA VARIÁVEL RESPOSTA .....	32
4.4.1. <i>Transformação logarítmica</i> .....	32
4.4.2. <i>Transformação Box-Cox</i> .....	33

4.4.3. <i>Transformação de Johnson</i> .....	34
4.5. REGRESSÃO LINEAR MÚLTIPLA COM VARIÁVEL TRANSFORMADA .....	35
4.5.1. <i>Avaliação de multicolinearidade</i> .....	37
4.5.2. <i>Resultado da regressão linear múltipla com transformada</i> .....	38
4.6. REGRESSÃO LINEAR MÚLTIPLA CONFORME LITERATURA .....	39
4.6.1. <i>Avaliação de multicolinearidade</i> .....	40
4.6.2. <i>Resultado da regressão linear múltipla com transformada</i> .....	41
<b>5. CONCLUSÃO</b> .....	<b>43</b>
<b>REFERÊNCIAS</b> .....	<b>46</b>

## 1. INTRODUÇÃO

Entende-se sismicidade como sinônimo de atividade sísmica, ocorrência de terremotos ou vibrações na terra (BATES; JACKSON, 1984). Assim sendo, a sismicidade é parte fundamental do nosso planeta, participando de sua gênese e seu desenvolvimento, bem como apresentando importante papel na vida humana (PRESS et al., 2006).

Na indústria da mineração, particularmente, as vibrações têm papel fundamental, sendo consequência direta dos desmontes de rocha imprescindíveis para a extração mineral. Para além disso, muitas vezes o contexto geológico dos corpos minerais são complexos em termos tectônicos, apresentando desafios na gestão segura das estruturas face aos esforços dinâmicos que as vibrações impõem sobre as estruturas de engenharia (KRAMER, 1996). Nesse cenário, o entendimento sobre eventos sísmicos e vibrações, ou *ground motions*, mostra-se bastante conveniente aos engenheiros, na medida em que permite prover critérios de projeto mais próximos à realidade, o que culmina em prover, no limite, segurança à sociedade.

Nesse contexto, não há confiabilidade na previsão de terremotos, nem tampouco seus efeitos, em virtude do alto grau de variabilidade dos geomateriais, bem como do ainda escasso conhecimento do interior da Terra. Não obstante, já há base de dados suficiente para se estabelecer modelos de movimentação de terreno (*ground motion models*, GMM). Os GMM são equações que associam potência e localização dos eventos sísmicos às vibrações que eles causam em pontos de interesse, com base em banco de dados coletados por estações sismológicas (MENDECKI, 2016).

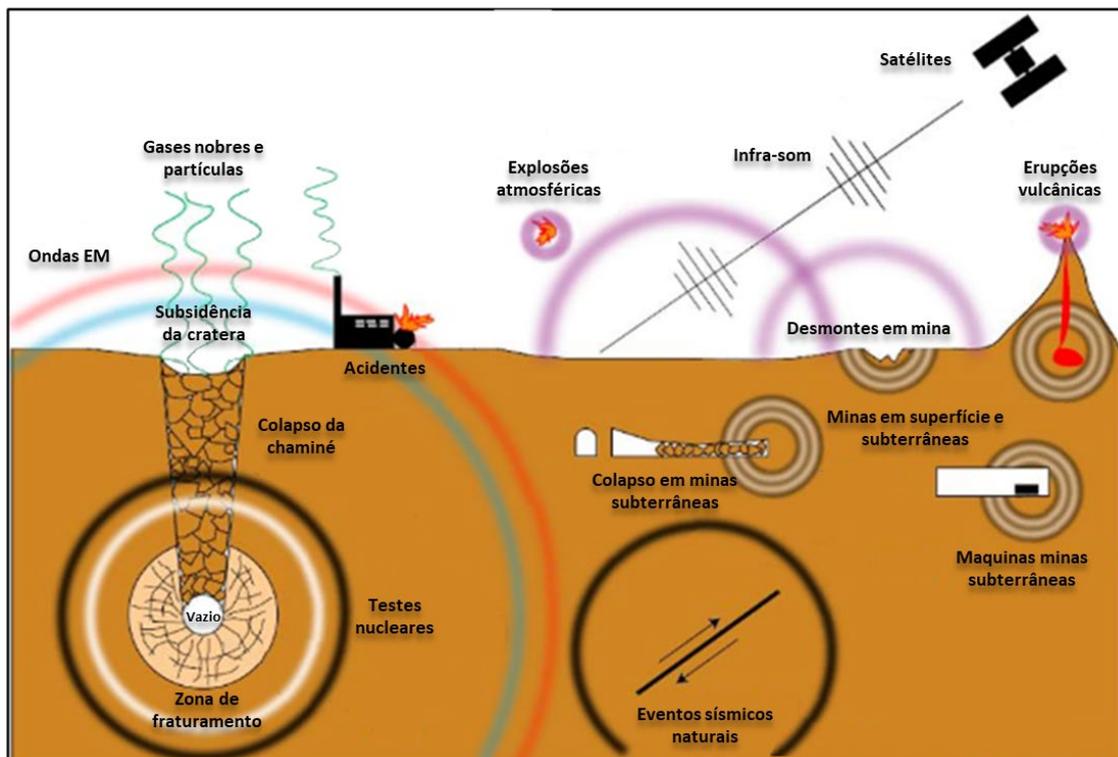
É com base no conceito de GMM que o presente trabalho se constrói, buscando propor uma regressão que associe estatisticamente os fenômenos que os sismólogos já associam empiricamente. O objetivo é, portanto, avaliar, com base no banco de dados da uma mina de fosfato do interior de São Paulo, Brasil, a força estatística de uma regressão que relacione os fenômenos usualmente empregados no GMM, já consagrada nos estudos sismológicos.

## 2. CONTEXTUALIZAÇÃO TEÓRICA

O estudo da sismologia é imprescindível para boas práticas de engenharia, uma vez que permite a obtenção de parâmetros a serem considerados em projetos das mais diferentes estruturas, sem os quais a habitação em regiões como Califórnia, Japão e Vancouver seria inviável.

Fontes sísmicas são fontes pontuais que geram ondas muito mais longas que seus próprios tamanhos lineares (GRECHKA & HEIGL, 2017). Diversas atividades são potenciais fontes sísmicas, destacando-se atividade de falhas geológicas, liberação de gases como metano, detonações superficiais ou subterrâneas, tráfego de equipamentos e desmontes de mina. A Figura 2.1 exemplifica algumas fontes sísmicas.

Figura 2.1 – Esquema de diferentes tipos de fontes sísmicas.



Fonte: Adaptado de Seissol (2014).

As diferentes fontes sísmicas produzem ondas que se propagam no meio e o deformam. Tais deformações são chamadas de “*ground motion*”, ou movimentação de terreno, e são vulgarmente conhecidas como vibração. A partir da análise das

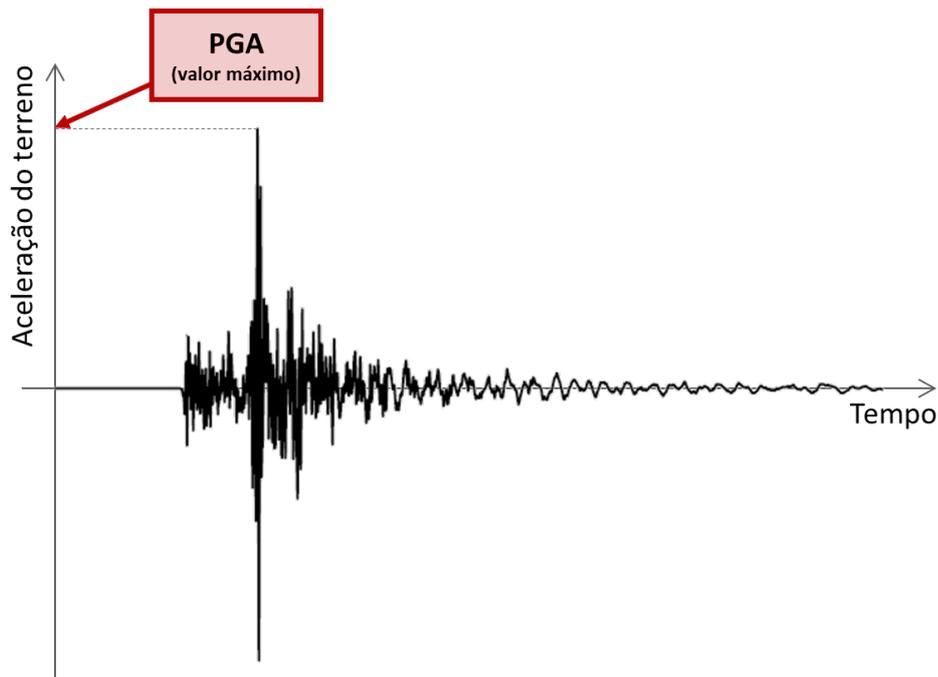
ondas produzidas pelos sismos, extrai-se a maioria das informações acerca desses eventos (STEIN & WYSESSION, 2003).

Para o presente estudo, três informações são importantes, detalhadas a seguir.

### 2.1. ACELERAÇÃO DE PICO DE TERRENO [M/S<sup>2</sup>]

Comumente chamado de PGA, em virtude do nome em inglês (*peak ground acceleration*), indica o pico de aceleração do terreno (aceleração máxima do solo) durante um evento. É obtido diretamente a partir do sismograma, sendo um dado auferido pelo geofone. A Figura 2.2, a seguir, indica a identificação do PGA a partir de um sismograma típico.

Figura 2.2 – Esquema de sismograma típico com apontamento do PGA.



Fonte: Elaborado pela autora.

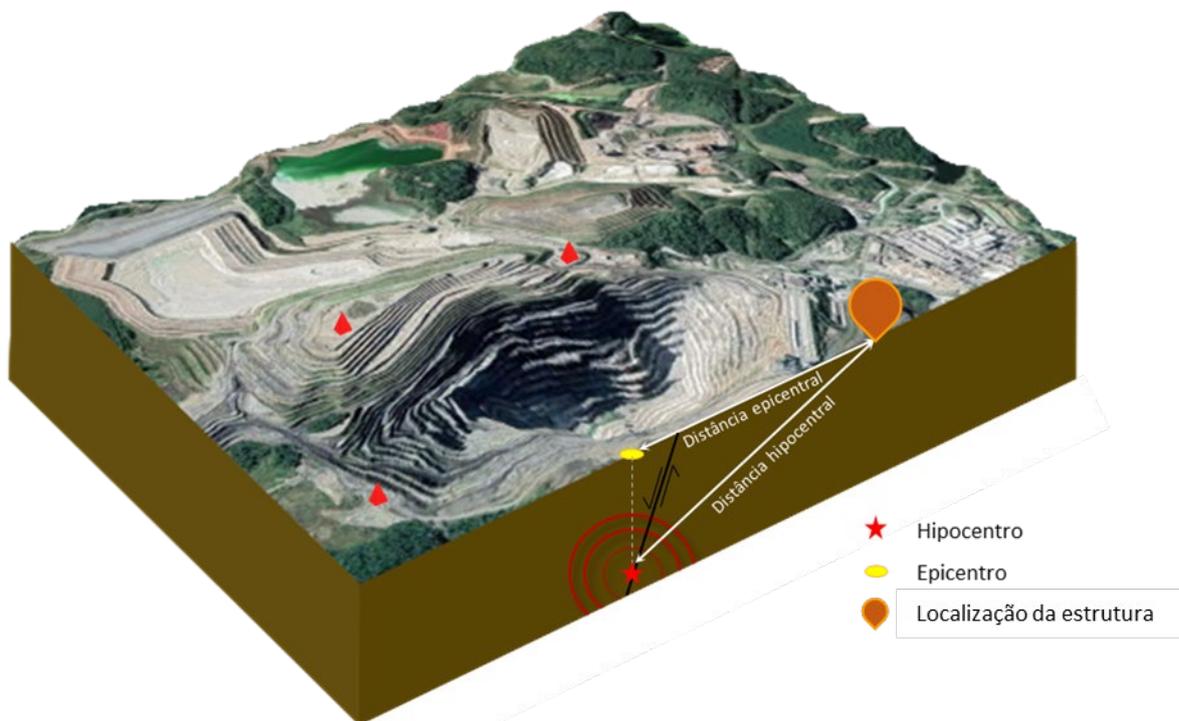
### 2.2. POTÊNCIA SÍSMICA [M<sup>3</sup>]

Diferentemente do PGA, a potência sísmica não é diretamente medida pelo sensor e, portanto, é calculada a partir de outros parâmetros. Para tal, é necessário mais de um registro (sismogramas de diferentes sensores). É, portanto, um dado indireto, porém derivado de maneira confiável (MENDECKI, 2013). Essa grandeza está associada ao tamanho do evento sísmico, ou seja, à energia que ele libera.

### 2.3. DISTÂNCIA [M]

A distância, por fim, é a extensão linear entre a fonte do evento sísmico (epicentro) ao objeto de interesse – no presente caso, a cava. A ela dá-se o nome de distância epicentral, conforme indicado na Figura 2.3. Portanto, é uma projeção superficial da fonte sísmica até o ponto no qual deseja-se estimar a vibração oriunda de tal evento, normalmente onde há uma estrutura de engenharia, ou estudos para sua implantação.

Figura 2.3 – Ilustração da distância epicentral, adotada no presente trabalho.



Fonte: Elaborado pela autora.

### 2.4. MODELOS DE MOVIMENTAÇÃO DE TERRENO

Conhecidos como GMM, em virtude do nome em inglês (*ground motion models*), trata-se de equações de previsão simples para a aceleração de pico do terreno, a partir de um evento de potência e distância conhecidas (MENDECKI, 2013). São, portanto, relações de atenuação sísmica do meio em virtude do evento original, apontado pela potência sísmica, de acordo com algum ponto de interesse, imputado na equação a partir da distância.

Há inúmeros modelos de movimentação de terreno (PEERC, 2011; Douglas, 2022), com diferentes soluções estatísticas. Não obstante, é notória a ocorrência de modelos lineares (MENDECKI, 2016), na forma:

$$PGA = c_1P + c_2(d) \quad \text{Eq. 1}$$

Na qual: *PGA*: aceleração de pico do terreno;  
*P*: potência sísmica (ou magnitude);  
*d*: distância entre a fonte sísmica e o ponto de interesse; e  
*c*<sub>1</sub> e *c*<sub>2</sub>: coeficientes obtidos a partir da regressão.

Ainda mais frequentes que modelos lineares, são aqueles que consideram as variáveis *PGA* e distância com transformações logarítmicas (Denham & Small, 1971; Denham et al., 1973; Ambraseys, 1975a; 1975b; 1978a; Faccioli, 1979; Costa et al., 1998; Ólafsson & Sigbjörnsson, 1999; Alarcón, 2003; Ghodrati Amiri et al., 2010). A esses modelos, dá-se a seguinte formulação, adotando-se os símbolos supracitados:

$$\log (PGA) = c_1P + c_2 \cdot \log (d) \quad \text{Eq. 2}$$

Modelos mais elaborados foram também propostos, porém requerem um banco de dados bastante robusto e, sobretudo, filtrado em virtude do tipo de evento sísmico de origem (PEERC, 2011; Douglas, 2022). Nesse sentido, embora vanguarda nessa ciência, não serão contemplados no presente estudo.

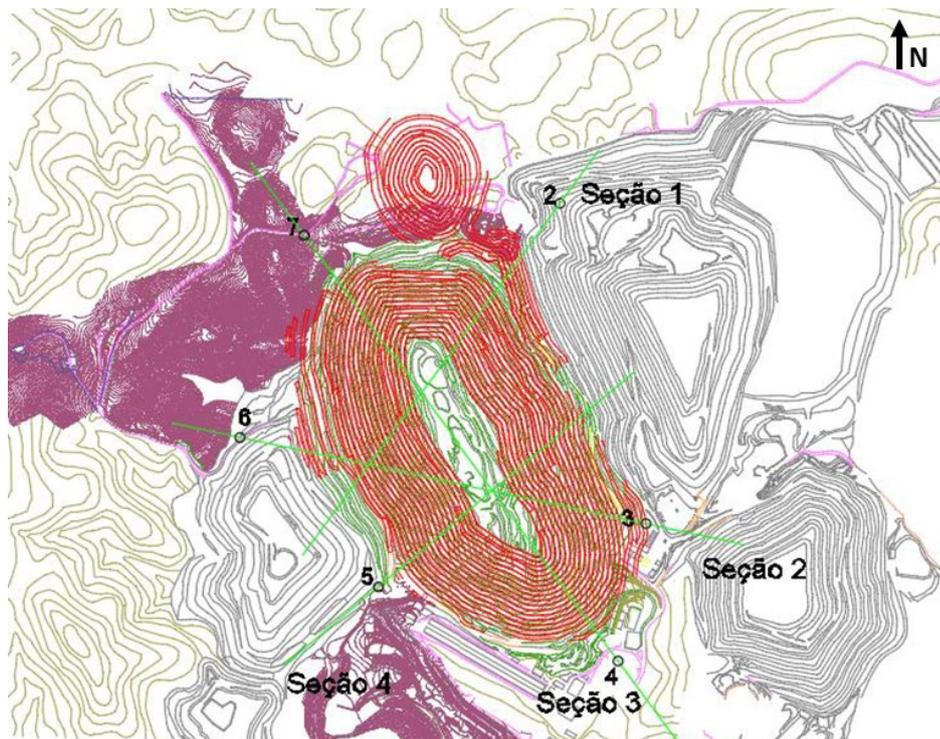
### 3. MÉTODOS

#### 3.1. COLETA E SELEÇÃO DE DADOS

Os dados do presente estudo são originados de um sistema de monitoramento microsísmico instalado em uma cava de fosfato, no interior de São Paulo, Brasil. Trata-se de geofones instalados em diferentes profundidades na cava, com a finalidade de se monitorar as vibrações de diferentes fontes, como eventos naturais (terremotos), detonações e ruptura de taludes.

A Figura 3.1, a seguir, apresenta a localização dos geofones, instrumentos empregados para monitoramento dos *ground motions*, na cava de interesse.

Figura 3.1 – Localização dos geofones ao redor da cava monitorada.



Fonte: Adaptado de MLF Geomecânica (2017).

O monitoramento é composto por doze geofones, que monitoramento constantemente as vibrações no meio. Quando engatilhados, ou seja, quando 3 ou mais são disparados em período inferior a 10 segundos, configuram, no sistema, um evento sísmico, seja ele natural ou não.

### **3.2. BANCO DE DADOS EMPREGADO**

O período selecionado para o banco de dados compreende de primeiro de maio de 2018 a trinta de abril de 2019. Foram considerados eventos aceitos pela empresa responsável pelo processamento sísmico dos dados, o que implica em empregar apenas dados validados pela empresa, de forma a terem sido descartados dados entendidos como “espúrios”, que se apresentam como ruídos sísmicos, não permitindo confiável localização da fonte e estimativa de potência sísmica e, conseqüentemente, incorrendo em importantes erros no banco de dados para a regressão a ser realizada.

Dessa forma, foram considerados 18.491 registros, que são compostos por uma trinca de dados de aceleração de pico de terreno (PGA, do inglês *peak ground acceleration*) [ $m/s^2$ ], potência sísmica [ $m^3$ ] e distância [m] entre a fonte sísmica e o ponto de registro da vibração, variáveis amplamente consolidadas na sismologia para elaboração de GMM.

### **3.3. SELEÇÃO DE VARIÁVEIS**

Face ao objetivo do presente estudo, as variáveis estavam postas, uma vez que, na literatura, os modelos de movimento do terreno (GMM) apresentam aceleração de pico do terreno (PGA) em função da potência sísmica e da distância entre a fonte sísmica e o ponto de interesse. Raras as vezes é possível encontrar a magnitude do evento ao invés da potência sísmica (MENDECKI, 2016). No entanto, o cálculo de magnitude do evento é menos seguro e, portanto, incorre em mais erros do que o cálculo de potência sísmica (MENDECKI, 1997; GRECHKA; HEIGL, 2017b), motivo pelo qual optou-se por prosseguir com a potência sísmica como variável independente, juntamente à distância.

Destaca-se, todavia, que não se espera que as variáveis potência sísmica e distância sejam plenamente capazes de explicar toda a variabilidade da aceleração de pico do terreno. A propagação de ondas mecânicas no meio é um fenômeno bastante complexo, com modelagem satisfatória apenas recente, e que envolve alto grau de incerteza, sobretudo por estar associado a diversos geomateriais, com diferentes propriedades e, ainda assim, pouco homogêneos ou isotrópicos.

## 4. RESULTADOS

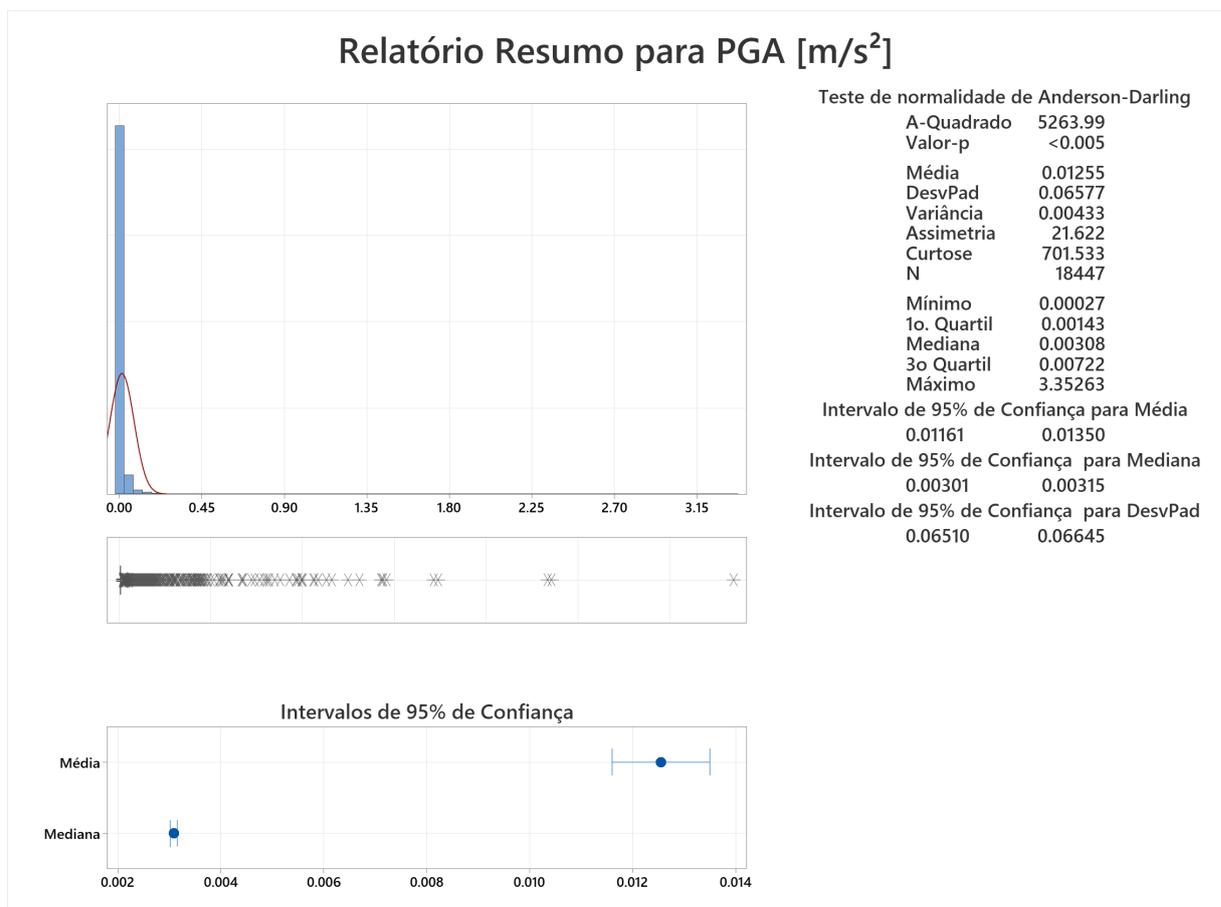
### 4.1. ANÁLISE EXPLORATÓRIA

Nesse subcapítulo, as variáveis distância, potência sísmica e PGA, disponíveis no banco de dados, serão individualmente observadas, a fim de obter informações básicas para compreender o comportamento de cada uma.

#### 4.1.1. Aceleração de pico do terreno

Na Figura 4.1 estão disponíveis estatísticas descritivas da variável independente “PGA”, dada em metros por segundo ao quadrado ( $m/s^2$ ), a seguir. Ressalta-se que é uma variável sempre maior que zero.

Figura 4.1 – Resumo estatístico da variável independente “PGA”.

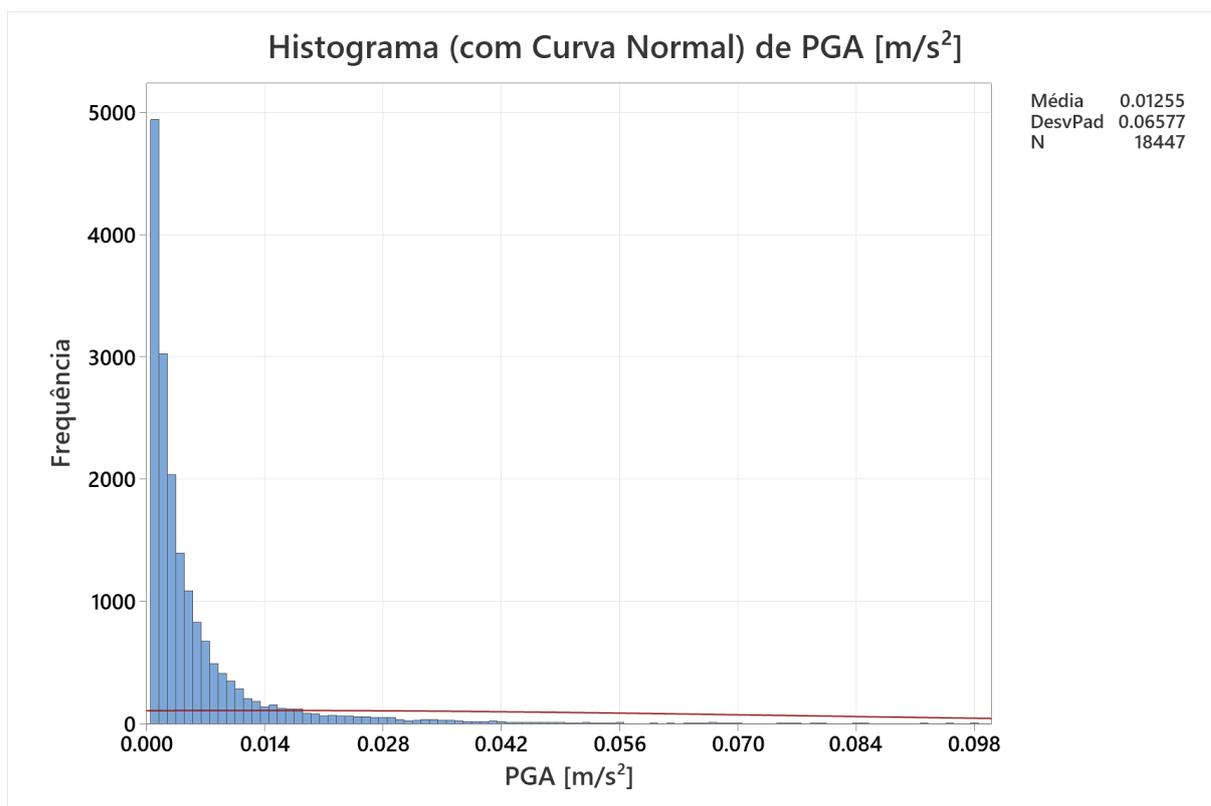


Fonte: Elaborado pela autora.

Observa-se, de antemão, o forte caráter assimétrico da distribuição. Tal fato é esperado pela sismologia clássica, que espera que vibrações pequenas sejam muito mais frequentes, culminando em uma distribuição bastante assimétrica à esquerda, como observado. O histograma da variável também corrobora essa questão. Observa-se que os valores variam de 0,00027 m/s<sup>2</sup> a 3,35263 m/s<sup>2</sup>, apresentando distribuição unimodal com assimetria bastante positiva, com valor médio de 0,01255 m/s<sup>2</sup> e mediana de 0,00308 m/s<sup>2</sup>.

Observa-se a forte influência de outliers, que não permitem avaliar com clareza o principal comportamento na maior parte da distribuição. Por isso, foi elaborado outro histograma, desconsiderando-se valores superiores a 0.10 para fins de visualização do comportamento na maior parte dos indivíduos, como apontado na Figura 4.2.

Figura 4.2 – Histograma da variável dependente “PGA” com truncamento superior para melhor visualização de maior parte dos indivíduos.

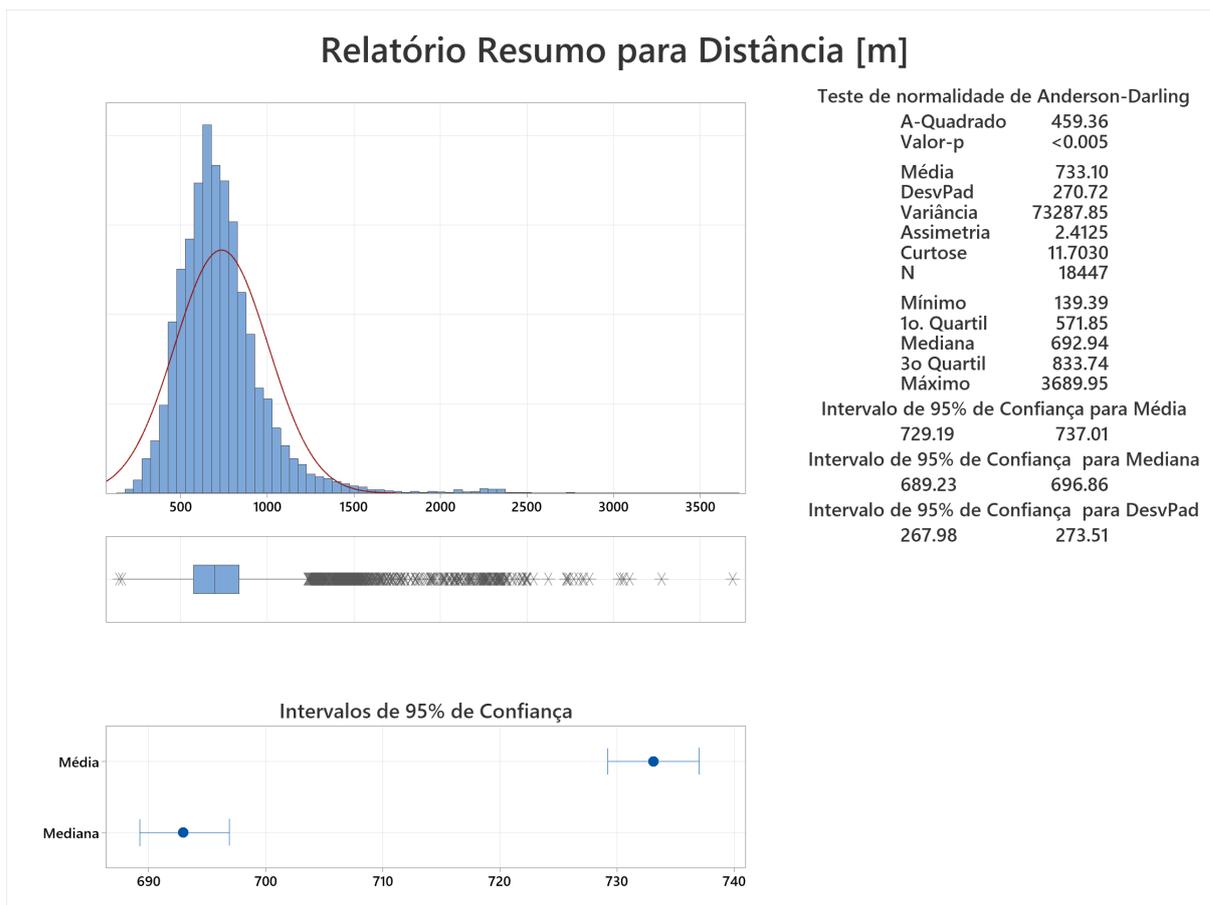


Fonte: Elaborado pela autora.

#### 4.1.2. Distância

Na Figura 4.3 estão disponíveis estatísticas descritivas da variável independente “distância”, dada em metros (m), a seguir. Ressalta-se que é uma variável sempre maior que zero.

Figura 4.3 – Resumo estatístico da variável independente “distância”.



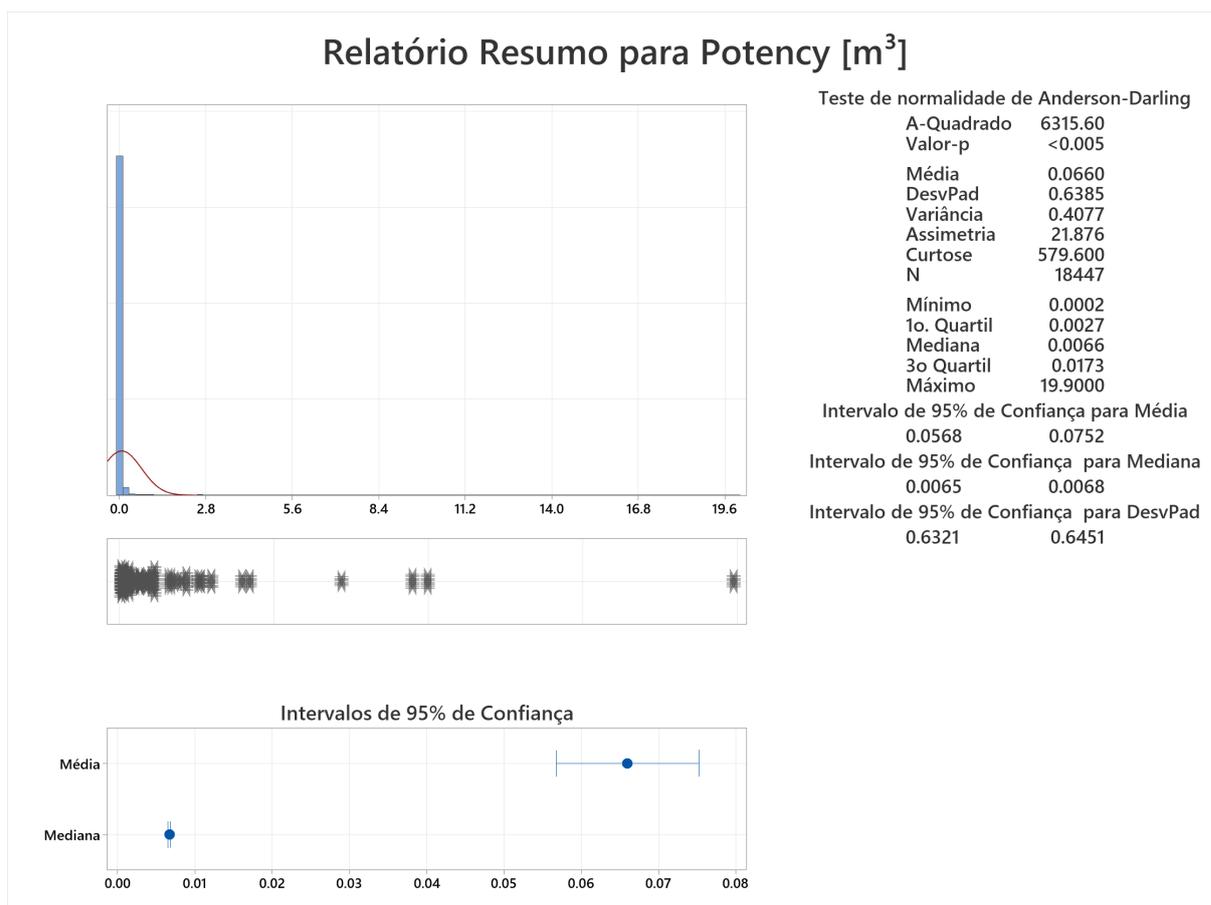
Fonte: Elaborado pela autora.

Por ser a média altamente influenciável por outliers, sabe-se, pelos resultados, que a distância é uma variável com assimetria à esquerda, sendo a média maior que a mediana. O histograma da distribuição corrobora as conclusões dos parâmetros identificados. Observa-se que os valores variam de 139,39 m a 3689,95 m, apresentando distribuição unimodal com assimetria positiva, com valor médio de 733,10 m e mediana de 692,94 m.

### 4.1.3. Potência sísmica

Na Figura 4.4 estão disponíveis estatísticas descritivas da variável independente “potência”, dada em metros cúbicos ( $m^3$ ), a seguir. Ressalta-se que é uma variável sempre maior que zero.

Figura 4.4 – Resumo estatístico da variável independente “potência”.

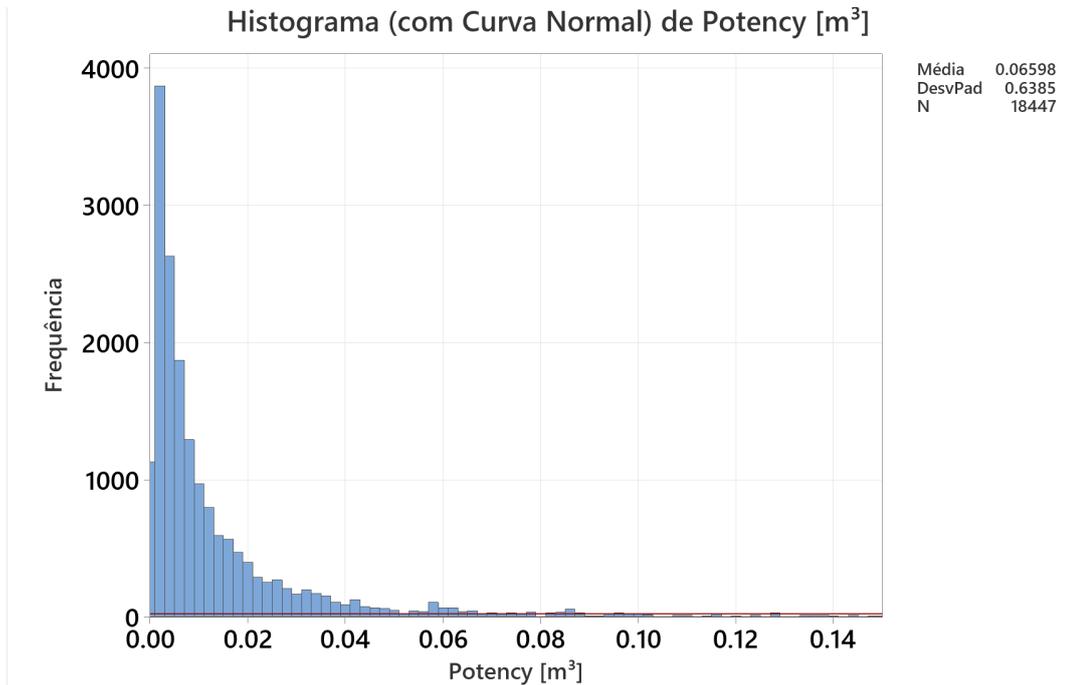


Fonte: Elaborado pela autora.

Observa-se que há assimetria à esquerda, a partir dos valores de mediana e média e do valor máximo observado. Observa-se que os valores variam de  $0,0002 m^3$  a  $19,9 m^3$ , apresentando distribuição unimodal com assimetria bastante positiva, com valor médio de  $0,066 m^3$  e mediana de  $0,007 m^3$ .

A fim de se melhor avaliar a maior parte dos indivíduos, a Figura 4.5 foi gerada, com truncamento em potência igual a  $0,15 m^3$ , com o exclusivo fim de melhorar a visualização ao redor da moda.

Figura 4.5 – Histograma da variável independente “potência sísmica” com truncamento superior para melhor visualização de maior parte dos indivíduos.

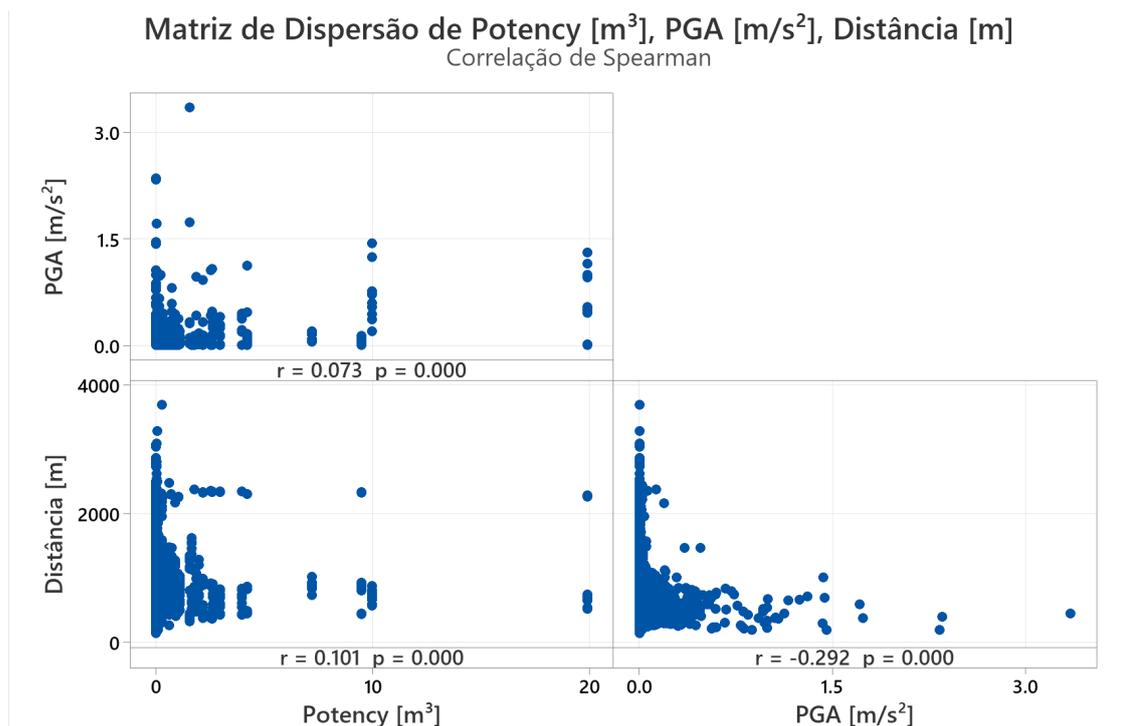


Fonte: Elaborado pela autora.

#### 4.1.4. Correlação entre as variáveis

Para se prosseguir com as regressões, foram avaliadas as correlações entre os três pares de variáveis, a fim de se descartar a colinearidade. Para tal, adotou-se a correlação de Spearman, que são baseados em classificações e são menos sensíveis à suposição de distribuição bivariada subjacente, quando comparados aos intervalos de confiança para a correlação de Pearson, conforme indicado na Figura 4.6.

Figura 4.6 – Matriz de dispersão pareada entre as três variáveis.



O gráfico de dispersão entre as variáveis apontou correlação estatisticamente significativa entre as duas variáveis, conforme ilustrado na Figura 4.6. Observa-se que o p-valor foi inferior ao nível de significância de interesse, de 0,05 (5%), não se rejeitando a hipótese nula, de correlação entre todos os pares de variáveis. A Figura 4.7 aponta os parâmetros obtidos na correlação de Spearman.

Figura 4.7 – Dados da matriz de dispersão pareada entre as três variáveis.

#### Correlações de Spearman pareadas

Amostra 1	Amostra 2	N	Correlação	IC de 95% para p	Valor-p
PGA [m/s <sup>2</sup> ]	Potency [m <sup>3</sup> ]	18447	0.073	(0.059, 0.088)	0.000
Distância [m]	Potency [m <sup>3</sup> ]	18447	0.101	(0.086, 0.115)	0.000
Distância [m]	PGA [m/s <sup>2</sup> ]	18447	-0.292	(-0.305, -0.278)	0.000

Fonte: Elaborado pela autora.

Observa-se, portanto, intervalos de confiança que não mudam a relação geral dos dados – (1) a relação entre PGA e potência e (2) a relação entre distância e potência são positivas (uma variável cresce na medida em que a outra aumenta). Por outro lado, (3) a relação entre distância e PGA têm correlação negativa – na medida em que a distância aumenta, o PGA tende a diminuir.

As relações 1 e 3 são esperadas, haja vista que o aumento da potência sísmica ou a proximidade do ponto de interesse da fonte sísmica (diminuição da distância) geram aumento na vibração registrada. A relação 2, por outro lado, não é óbvia, pois depende da geologia local, bem como do tipo de fonte sísmica, mas o resultado obtido é factível.

De toda forma, é compreensível que, mesmo havendo correlação entre as variáveis independentes, ambas sejam adotadas na literatura. A potência sísmica é uma variável dependente exclusivamente da fonte, ao passo que a distância é dependente da fonte, mas também do ponto de interesse. Assim, este indicador traz uma informação importante, associada à atenuação descrita pelo GMM, que não pode ser explicada pela potência sísmica.

#### **4.2. REGRESSÕES LINEARES SIMPLES**

A primeira avaliação de sensibilidade a se desempenha diz respeito a análise de regressões simples, em que se busque compreender a aceleração de pico do terreno (PGA) de acordo com apenas uma variável, seja distância ou potência sísmica.

Como regressão linear, os pressupostos aos quais o modelo deve respeitar são:

- Normalidade dos resíduos;
- Homocedasticidade da variância; e
- Relação de natureza linear entre as variáveis de interesse.

Caso a regressão culmine em resíduos e variâncias que não respeitem tais pressupostos, os testes de hipótese para aferir sua validade não são válidos, não sendo possível atestar a significatividade das variáveis.

##### **4.2.1. PGA versus distância**

Considerando-se o PGA como dependente exclusivamente da distância entre o evento sísmico e o ponto de interesse, tem-se o modelo de regressão linear simples como apresentado na Figura 4.8.

Figura 4.8 – Modelo de regressão linear simples para PGA predito por distância.

### Equação de Regressão

$$\text{PGA [m/s}^2\text{]} = 0.02885 - 0.000022 \text{ Distância [m]}$$

### Coeficientes

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	0.02885	0.00139	20.73	0.000	
Distância [m]	-0.000022	0.000002	-12.48	0.000	1.00

### Sumário do Modelo

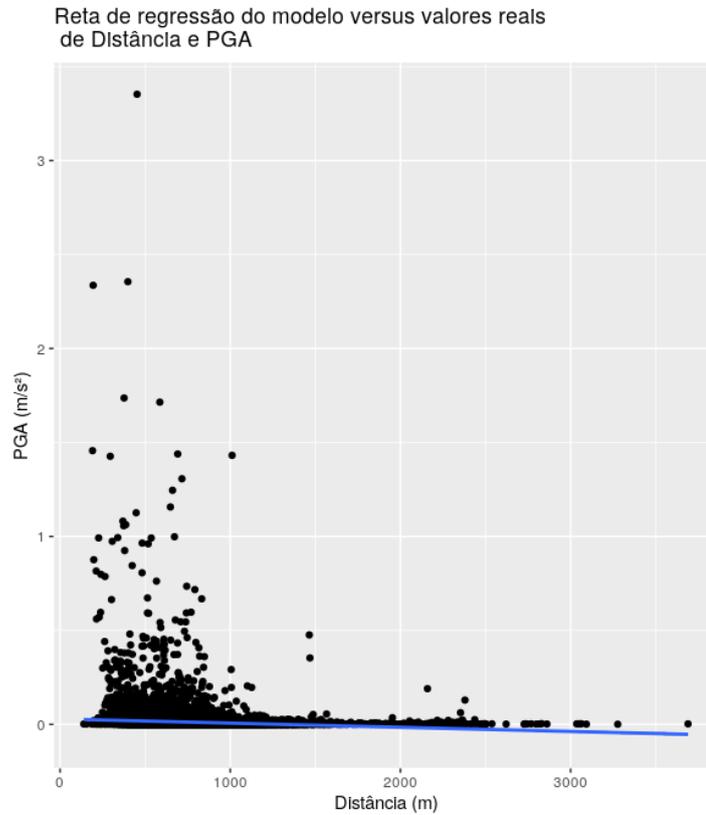
S	R2	R2(aj)	R2(pred)
0.0654933	0.84%	0.83%	0.81%

Fonte: Elaborado pela autora.

O coeficiente de determinação preditivo ( $R^2$  pred) do modelo foi de 0,81%, indicando que 99,19% da variabilidade do PGA não pode ser explicada pelo modelo. Nesse sentido, é um número insuficiente para considerarmos o modelo como adequado.

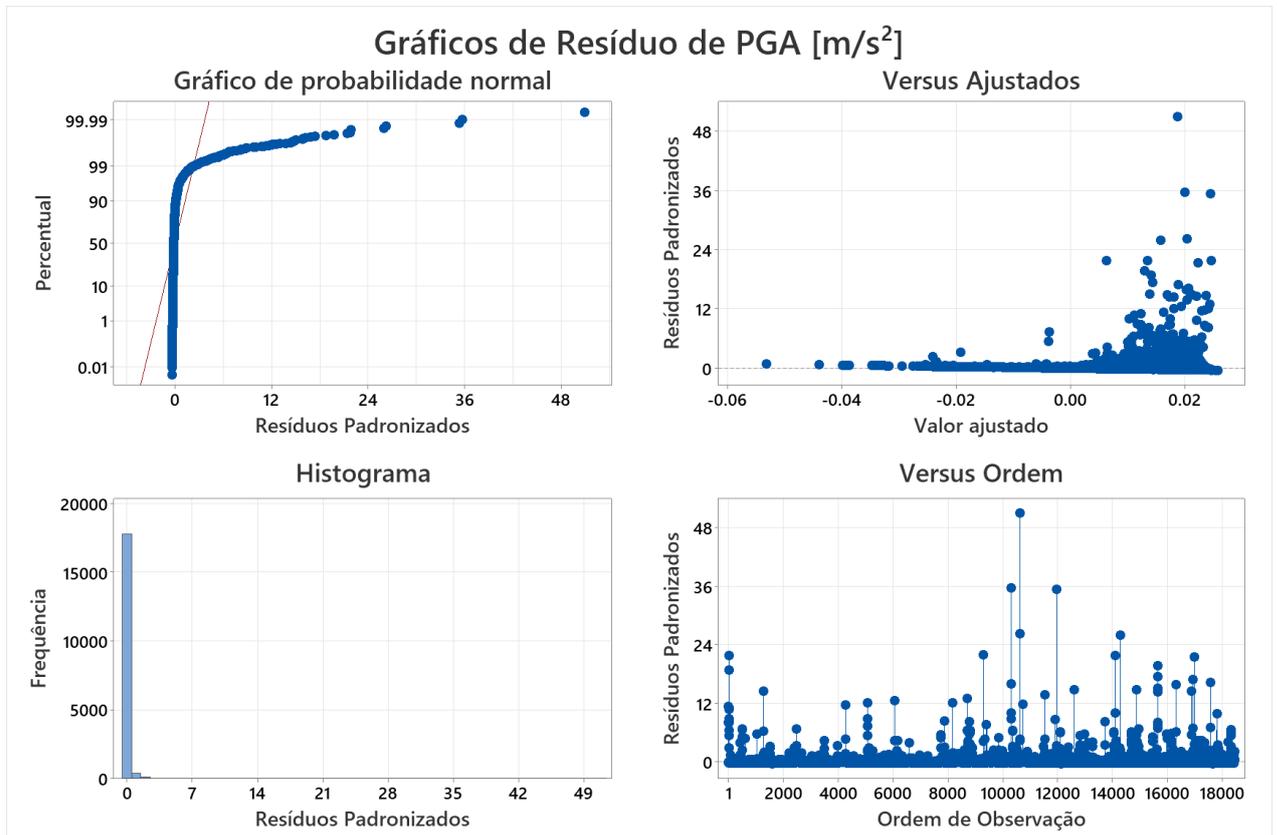
A Figura 4.9, a seguir, apresenta a reta de regressão linear para o modelo proposto, bem como os indivíduos presentes no banco de dados, evidenciando a divergência. A Figura 4.10, por sua vez, apresenta o gráfico de distribuição normal dos resíduos padronizados.

Figura 4.9 – Regressão simples entre PGA e distância para os dados considerados.



Fonte: Elaborado pela autora.

Figura 4.10 – Análise dos resíduos padronizados para o modelo de regressão linear simples PGA x distância.



Fonte: Elaborado pela autora.

Observa-se, portanto, que os pressupostos não são respeitados, de forma que o modelo não é válido. Não se trata de resíduos normalmente distribuídos (de acordo com o gráfico de probabilidade normal), tampouco de variância homocedástica (de acordo com os resíduos padronizados versus ajustados).

#### 4.2.2. PGA versus potência sísmica

Considerando-se o PGA como dependente exclusivamente da potência sísmica do evento, tem-se o modelo de regressão linear simples como apresentado na Figura 4.11.

Figura 4.11 – Modelo de regressão linear simples para PGA predito por potência.

#### Equação de Regressão

$$\text{PGA [m/s}^2\text{]} = 0.009964 + 0.039260 \text{ Potency [m}^3\text{]}$$

#### Coeficientes

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	0.009964	0.000450	22.14	0.000	
Potency [m <sup>3</sup> ]	0.039260	0.000701	56.00	0.000	1.00

#### Sumário do Modelo

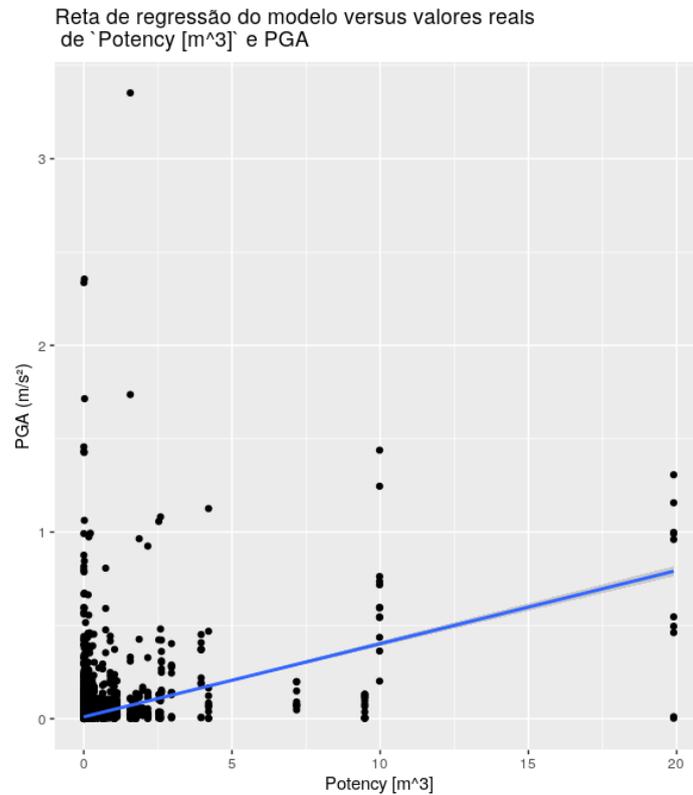
S	R2	R2(aj)	R2(pred)
0.0608037	14.53%	14.53%	14.09%

Fonte: Elaborado pela autora.

O coeficiente de determinação preditivo ( $R^2$  pred) do modelo foi de 14,09%, indicando que 85,91% da variabilidade do PGA não pode ser explicada pelo modelo. Ainda que bastante maior que o  $R^2$  preditivo do modelo considerando a distância como variável independente, esse valor ainda é baixo, não sendo considerado como adequado ao modelo proposto.

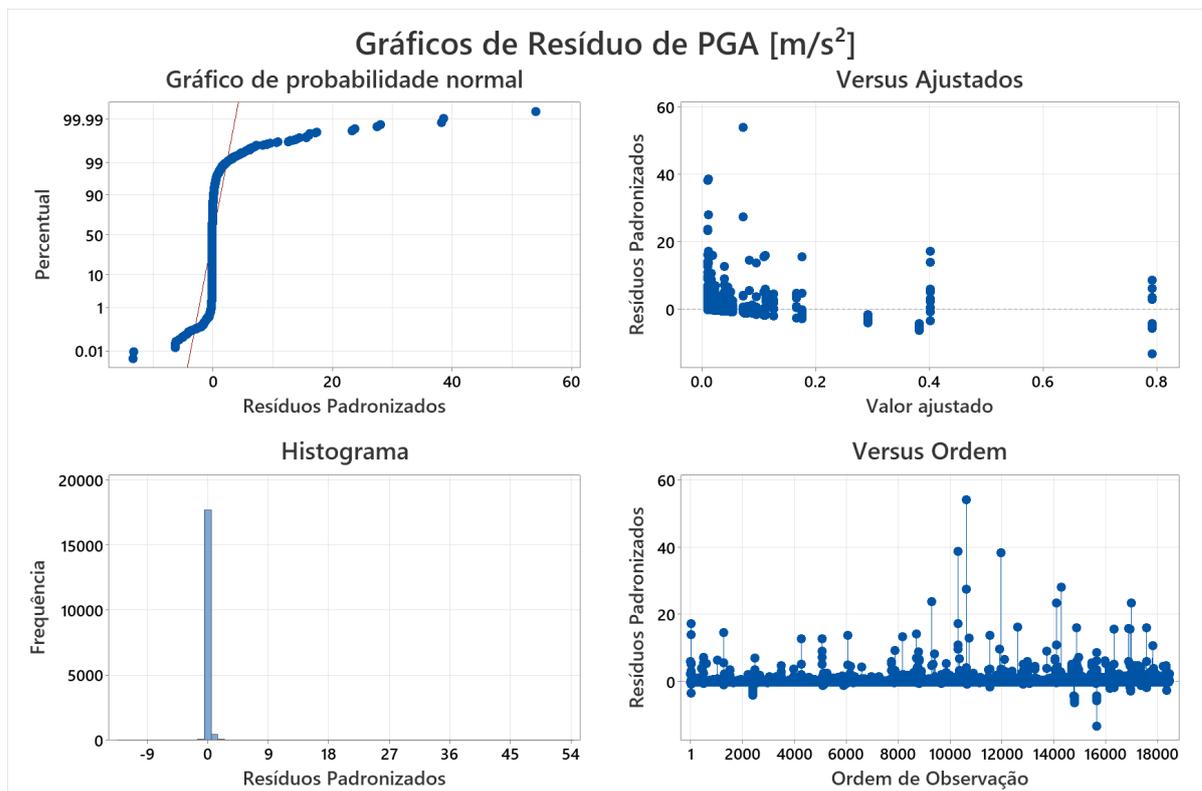
A Figura 4.12, a seguir, apresenta a reta de regressão linear para o modelo proposto, bem como os indivíduos presentes no banco de dados, evidenciando a divergência. A Figura 4.13, por sua vez, apresenta o gráfico de distribuição normal dos resíduos padronizados.

Figura 4.12 – Regressão simples entre PGA e potência sísmica para os dados considerados.



Fonte: Elaborado pela autora.

Figura 4.13 – Análise dos resíduos padronizados para o modelo de regressão linear simples PGA x potência sísmica.



Fonte: Elaborado pela autora.

Observa-se, portanto, que os pressupostos não são respeitados, de forma que o modelo não é válido. Não se trata de resíduos normalmente distribuídos (de acordo com o gráfico de probabilidade normal), tampouco de variância homocedástica (de acordo com os resíduos padronizados versus ajustados).

#### **4.3. REGRESSÃO LINEAR MÚLTIPLA**

Em virtude da não adequação dos modelos de regressão linear simples, foi tentada, também, a regressão linear múltipla, considerando-se as duas variáveis já consolidadas na literatura de sismologia.

##### *4.3.1. Avaliação de multicolinearidade*

Como supracitado, há correlação entre as variáveis independentes (distância e potência sísmica). Nesse sentido, há chance de multicolinearidade, ao se considerar a amostra de dados empregada.

##### *4.3.2. Resultado da regressão linear múltipla*

O primeiro modelo múltiplo a ser testado, naturalmente, será o modelo de regressão linear, em virtude de sua simplicidade. Há de se destacar, todavia, que os pressupostos desse modelo são:

- Normalidade dos resíduos;
- Homocedasticidade da variância; e
- Relação de natureza linear entre as variáveis de interesse.

Ou seja, os dados já não cumprem os pressupostos de antemão, haja vista que já se sabe que os dados não apresentam relação de natureza linear. Ainda assim, buscou-se avaliar a regressão linear a fim de se ter sensibilidade de seu coeficiente de determinação preditivo ( $R^2$  pred), que ajuda a determinar superajuste do modelo de regressão.

Dessa forma, o modelo de regressão linear obtido está exposto na Figura 4.14.

Figura 4.14 – Modelo de regressão linear simples para PGA predito tanto por distância, quanto por potência.

### Equação de Regressão

$$\text{PGA [m/s}^2\text{]} = 0.03144 + 0.040200 \text{ Potency [m}^3\text{]} - 0.000029 \text{ Distância [m]}$$

### Coefficientes

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	0.03144	0.00128	24.52	0.000	
Potency [m <sup>3</sup> ]	0.040200	0.000697	57.67	0.000	1.01
Distância [m]	-0.000029	0.000002	-17.87	0.000	1.01

### Sumário do Modelo

S	R2	R2(aj)	R2(pred)
0.0602857	15.98%	15.98%	15.54%

Fonte: Elaborado pela autora.

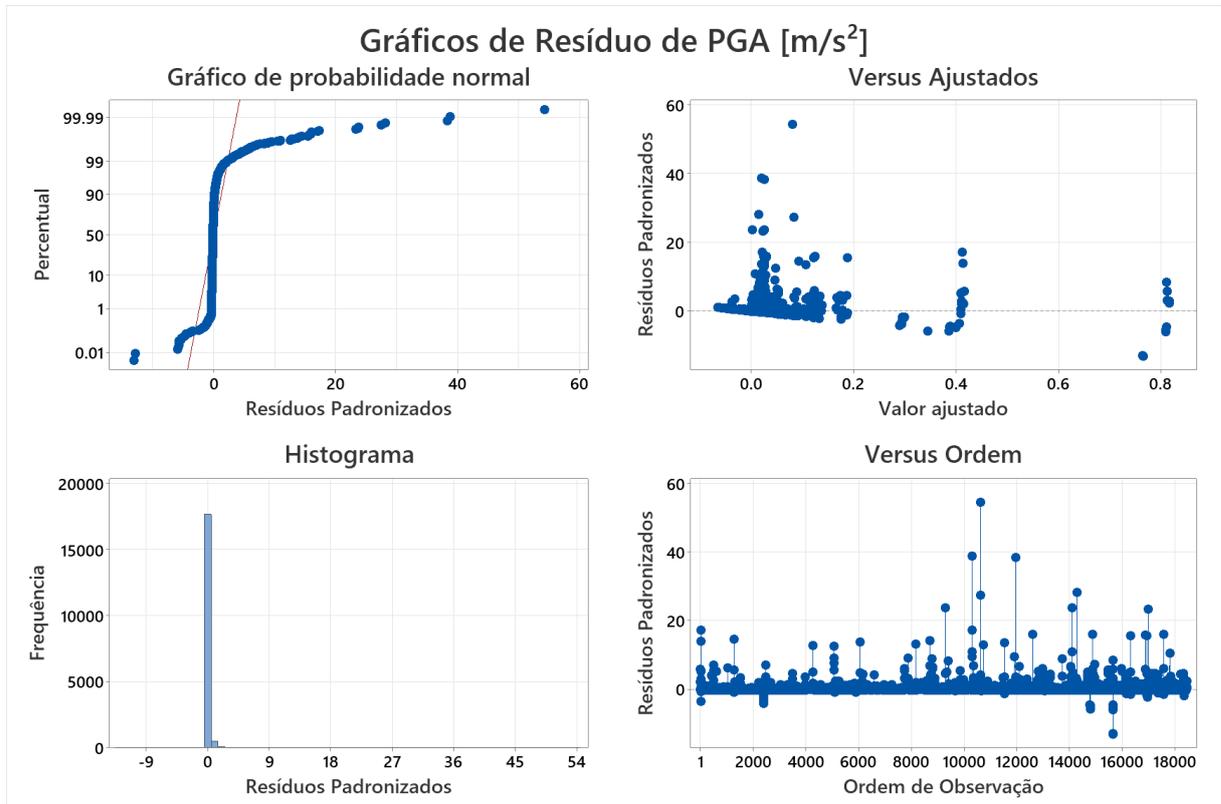
O coeficiente de determinação preditivo ( $R^2$  pred) do modelo foi de 15,54%, indicando que 84,46% da variabilidade do PGA não pode ser explicada pelo modelo. Esse valor é irrisoriamente maior que o modelo que considera a potência sísmica como única variável independente e, da mesma forma, é considerado inadequado ao modelo.

Pelo modelo, como esperado pela teoria de sismologia e pelas práticas dessa ciência nos cálculos de modelos de movimentação de terreno (GMM), as duas variáveis são importantes para ajudar a explicar a variável PGA, em virtude de terem apresentado grau de significância superior a 5% no teste.

O modelo fica, portanto, fraco estatisticamente e não se pode afirmar que consegue prever os dados corretamente. Tal observação é esperada, em virtude de o modelo não seguir os pressupostos estatísticos para sua validade.

A análise de resíduos é apresentada na Figura 4.15, a seguir.

Figura 4.15 – Análise dos resíduos padronizados para o modelo de regressão linear simples para PGA a partir de potência sísmica e distância.



Fonte: Elaborado pela autora.

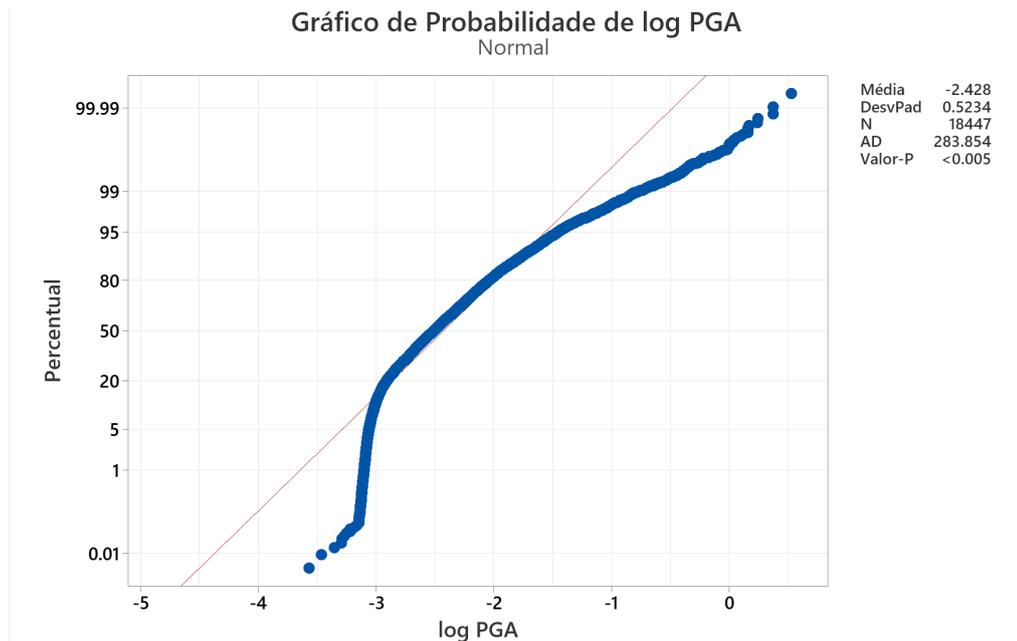
Observa-se, portanto, que os pressupostos não são respeitados, de forma que o modelo não é válido. Não se trata de resíduos normalmente distribuídos (de acordo com o gráfico de probabilidade normal), tampouco de variância homocedástica (de acordo com os resíduos padronizados versus ajustados).

#### 4.4. TRANSFORMAÇÃO DA VARIÁVEL RESPOSTA

##### 4.4.1. Transformação logarítmica

Um frequente artifício para a regressão quando não se aplica a linear é a transformação da variável resposta para tentar que o modelo linear se aplique, então, ao problema. Desta forma, em caráter de tentativa, foi realizada a transformação logarítmica, cujos dados foram dispostos no histograma apresentado na Figura 4.16, a seguir.

Figura 4.16 – Histograma da variável dependente “PGA” face a transformação logarítmica.



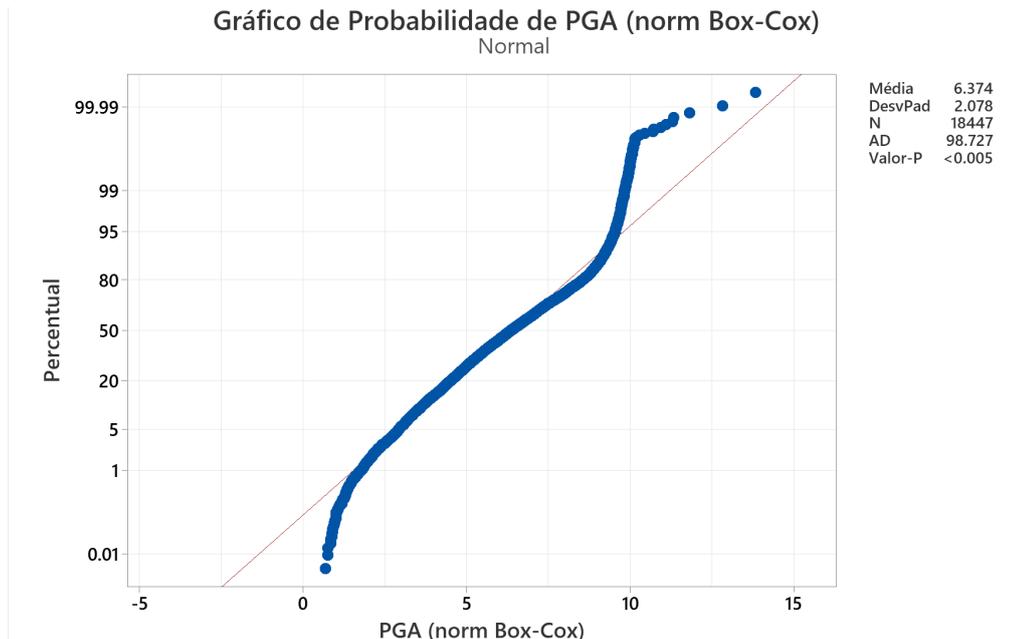
Fonte: Elaborado pela autora.

Ainda assim, observa-se que a variável não é normalmente distribuída, em virtude do baixo p-valor, descartando-se a hipótese nula, de que se trata de distribuição normal.

#### 4.4.2. Transformação Box-Cox

A transformação Box-Cox, adequada a dados positivos, também foi testada. O resultado do teste de normalidade Anderson-Darling está exposto na Figura 4.17. Observa-se que não foi suficiente para a normalização da variável resposta.

Figura 4.17 – Histograma da variável dependente “PGA” face a transformação Box-Cox.

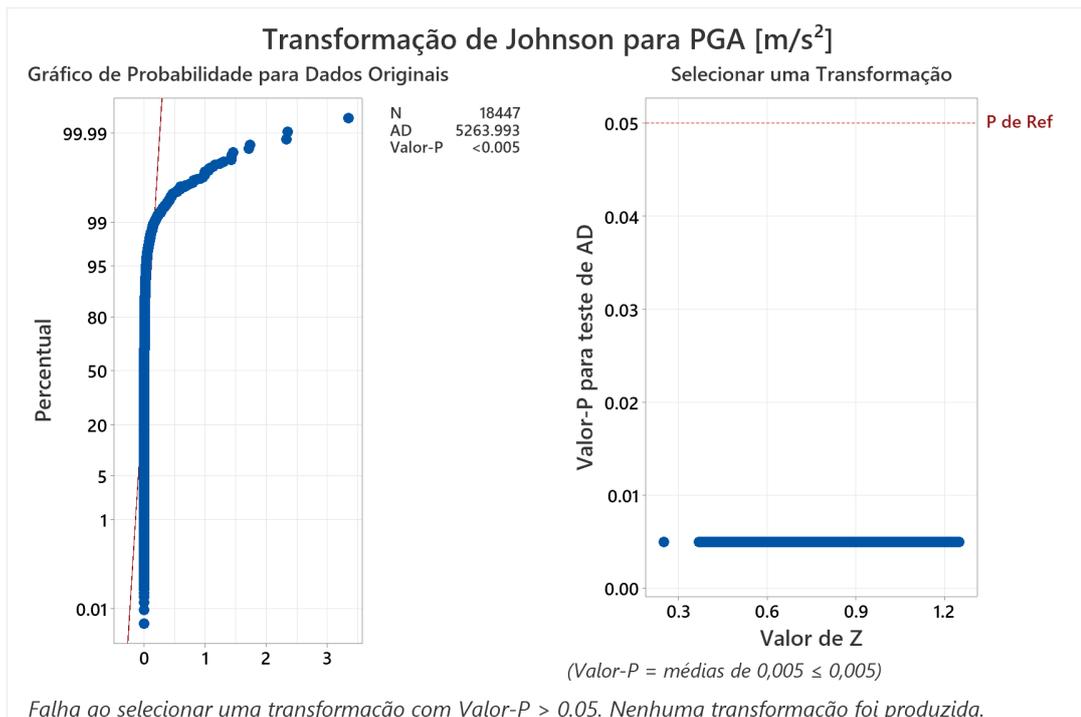


Fonte: Elaborado pela autora.

#### 4.4.3. Transformação de Johnson

A transformação de Johnson, por sua vez, adequada a dados positivos, também foi tentada (Figura 4.18). Porém, não foi possível convergir para transformação para nível de significância  $p$ -valor  $\geq 0,05$ . Portanto, a transformação não foi bem-sucedida.

Figura 4.18 – Histograma da variável dependente “PGA” face a transformação de Johnson.

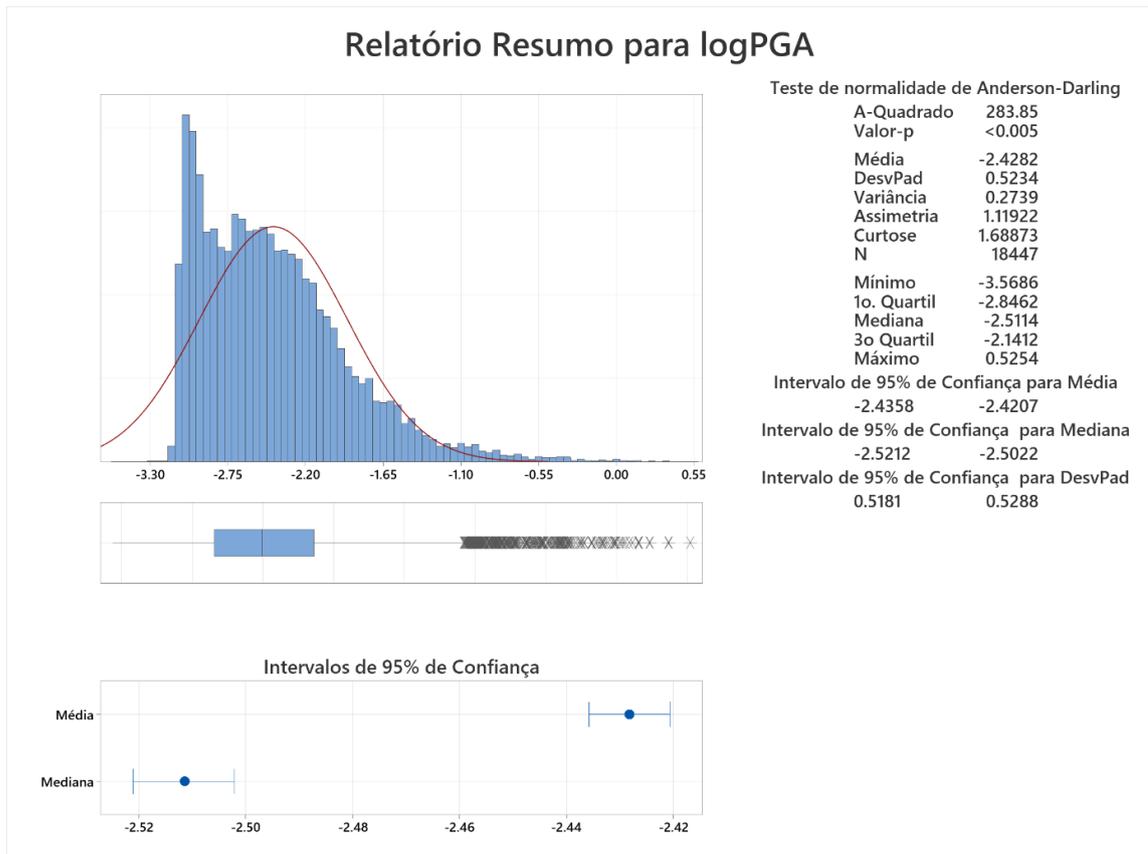


Fonte: Elaborado pela autora.

#### 4.5. REGRESSÃO LINEAR MÚLTIPLA COM VARIÁVEL TRANSFORMADA

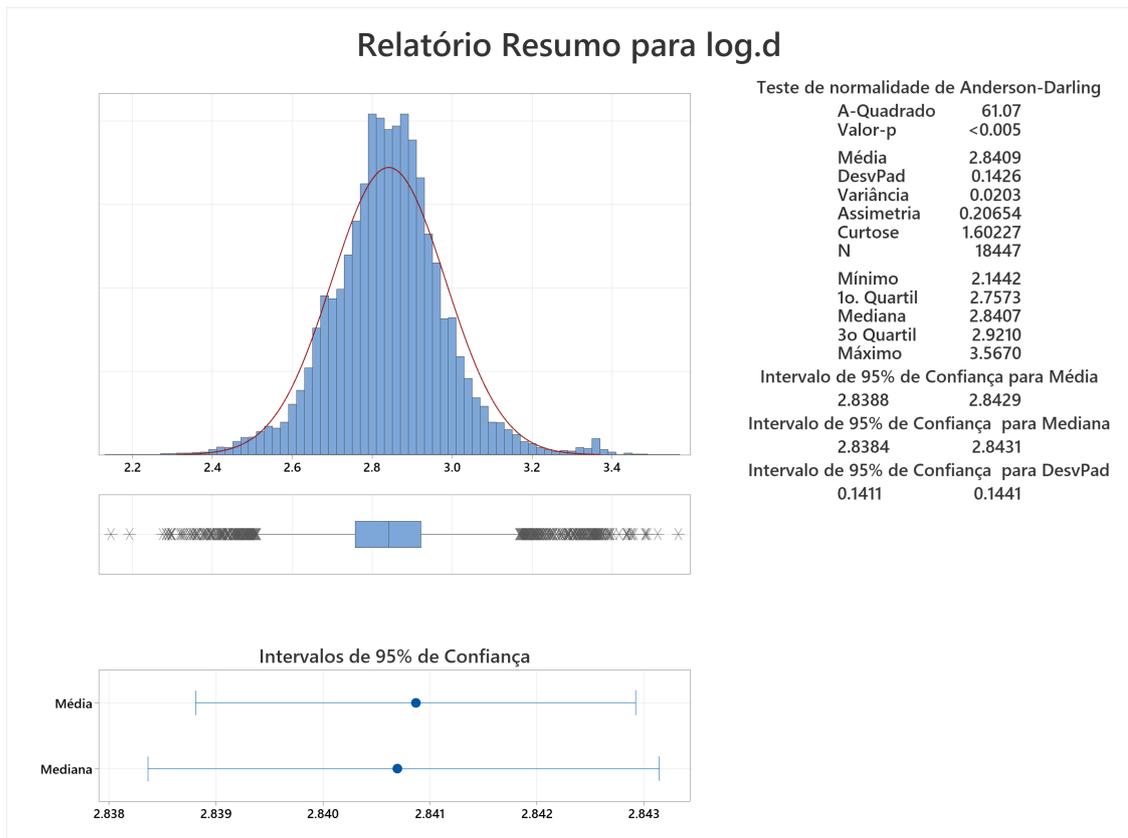
Em virtude das formas comumente adotadas nos modelos de movimentação de terreno (GMM), será adotado o modelo de logaritmo das três variáveis. Nesse sentido, os resumos estatísticos das variáveis transformadas estão apontados a seguir, da Figura 4.19 à Figura 4.21. A matriz de correlação está disponível na Figura 4.22.

Figura 4.19 – Resumo estatístico da variável independente “logPGA”.



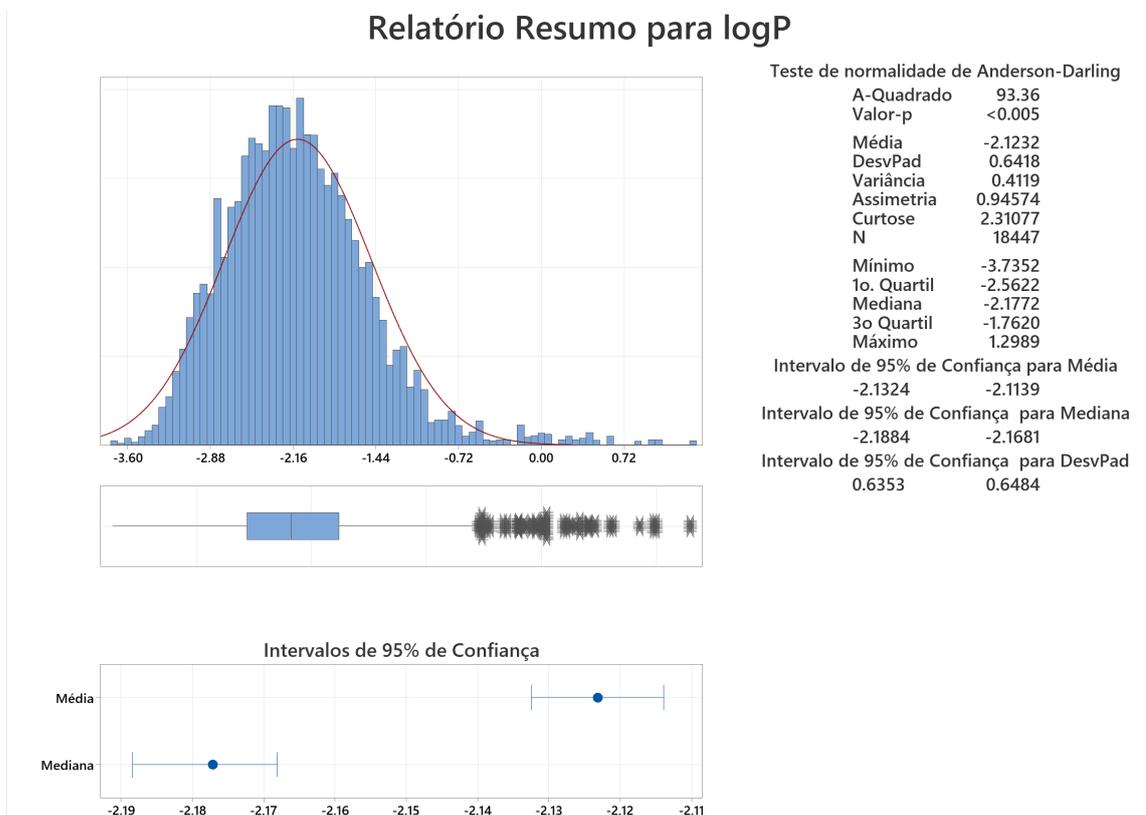
Fonte: Elaborado pela autora.

Figura 4.20 – Resumo estatístico da variável independente “log da distância”.



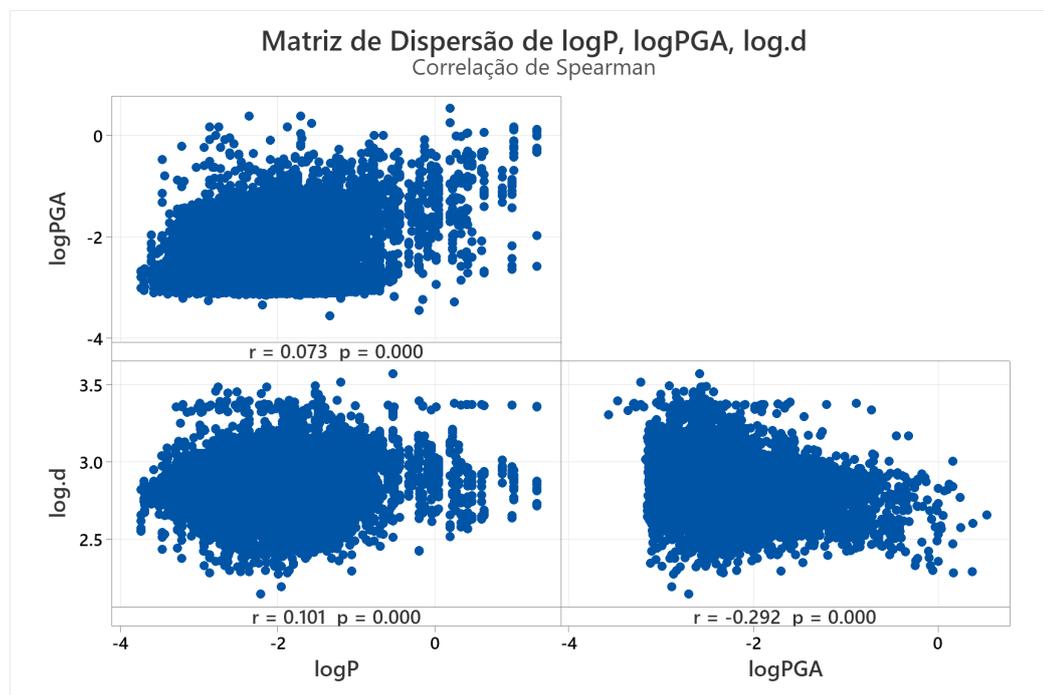
Fonte: Elaborado pela autora.

Figura 4.21 – Resumo estatístico da variável independente “log da potência”.



Fonte: Elaborado pela autora.

Figura 4.22 – Matriz de correlação, entre as três variáveis transformadas por algoritmo.



#### Correlações de Spearman pareadas

Amostra 1	Amostra 2	N	Correlação	IC de 95% para $\rho$	Valor-p
logPGA	logP	18447	0.073	(0.059, 0.088)	0.000
log.d	logP	18447	0.101	(0.086, 0.115)	0.000
log.d	logPGA	18447	-0.292	(-0.305, -0.278)	0.000

Fonte: Elaborado pela autora.

Observa-se, portanto, que as distribuições se tornaram menos assimétricas, se aproximando da normalidade. No tocante à matriz de correlação entre as três variáveis transformadas, os p-valores indicam que não se pode descartar a hipótese de correlação, com p-valores próximos a zero. Além disso, mantêm-se, conforme esperado, as tendências anteriores (relações positivas para PGA x potência e para potência x distância e relações negativas para PGA x distância).

#### 4.5.1. Avaliação de multicolinearidade

Como supracitado, há correlação entre as variáveis independentes (distância e potência sísmica). Nesse sentido, há chance de multicolinearidade, ao se considerar a amostra de dados empregada.

#### 4.5.2. Resultado da regressão linear múltipla com transformada

O modelo múltiplo empregado, considerando-se as variáveis transformadas, é linear. Lembra-se que os pressupostos desse modelo são:

- Normalidade dos resíduos;
- Homocedasticidade da variância; e
- Relação de natureza linear entre as variáveis de interesse.

Ou seja, os dados já não cumprem os pressupostos de antemão, haja vista que já se sabe que os dados não apresentam relação de natureza linear. Ainda assim, buscou-se avaliar a regressão linear a fim de se ter sensibilidade de seu coeficiente de determinação preditivo ( $R^2$  pred), que indica quanto da variabilidade da variável resposta pode ser explicada pelo modelo em xeque, que é o comumente aplicado na Sismologia.

Dessa forma, o modelo de regressão linear múltiplo obtido está exposto na Figura 4.23.

Figura 4.23 – Modelo de regressão linear simples para log do PGA predito tanto por log da distância, quanto por log da potência.

#### Equação de Regressão

$$\log\text{PGA} = 1.2728 + 0.22586 \log P - 1.1340 \log.d$$

#### Coeficientes

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	1.2728	0.0739	17.23	0.000	
logP	0.22586	0.00558	40.48	0.000	1.02
log.d	-1.1340	0.0251	-45.15	0.000	1.02

#### Sumário do Modelo

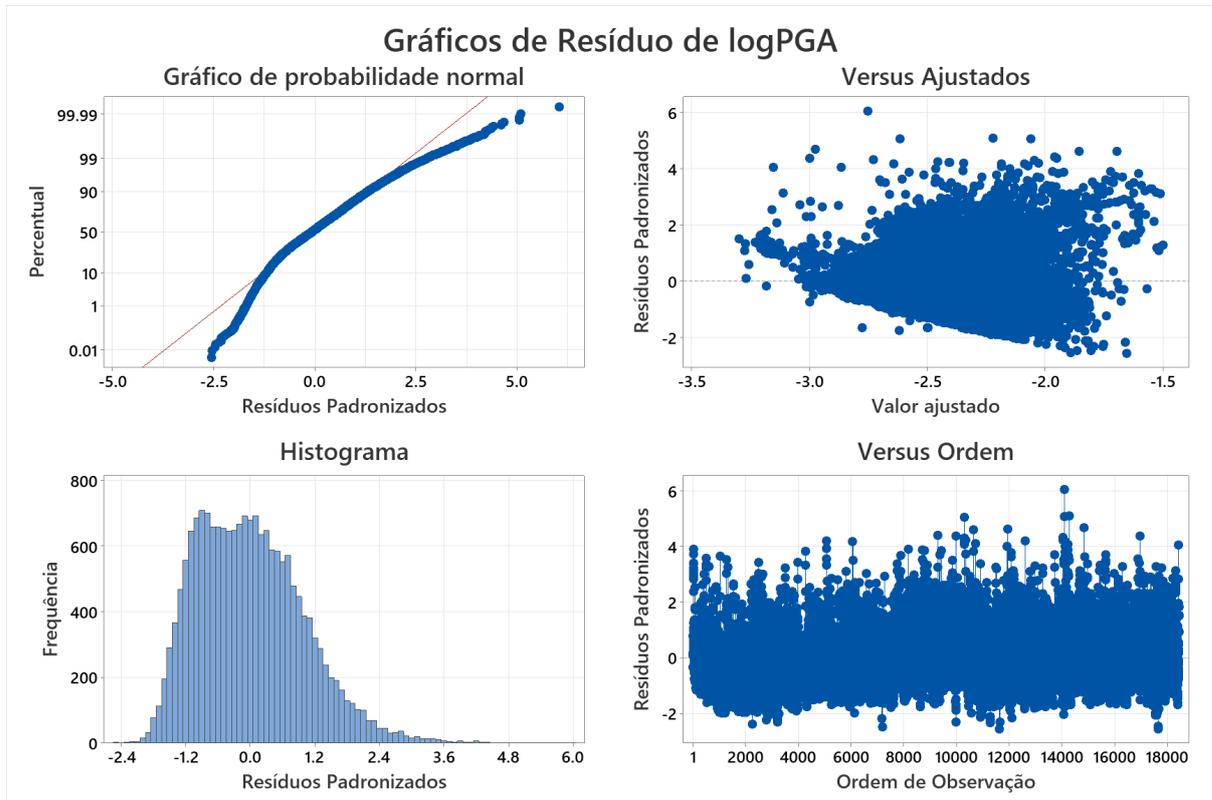
S	R2	R2(aj)	R2(pred)
0.482408	15.04%	15.04%	15.01%

Fonte: Elaborado pela autora.

O coeficiente de determinação preditivo ( $R^2$  pred) obtido foi de 15,01%, menor do que o modelo de regressão linear múltipla considerando-se as variáveis sem transformação.

A análise de resíduos é apresentada na Figura 4.24, a seguir.

Figura 4.24 – Análise dos resíduos padronizados para o modelo de regressão linear simples para log de PGA a partir de log da potência sísmica e log da distância.



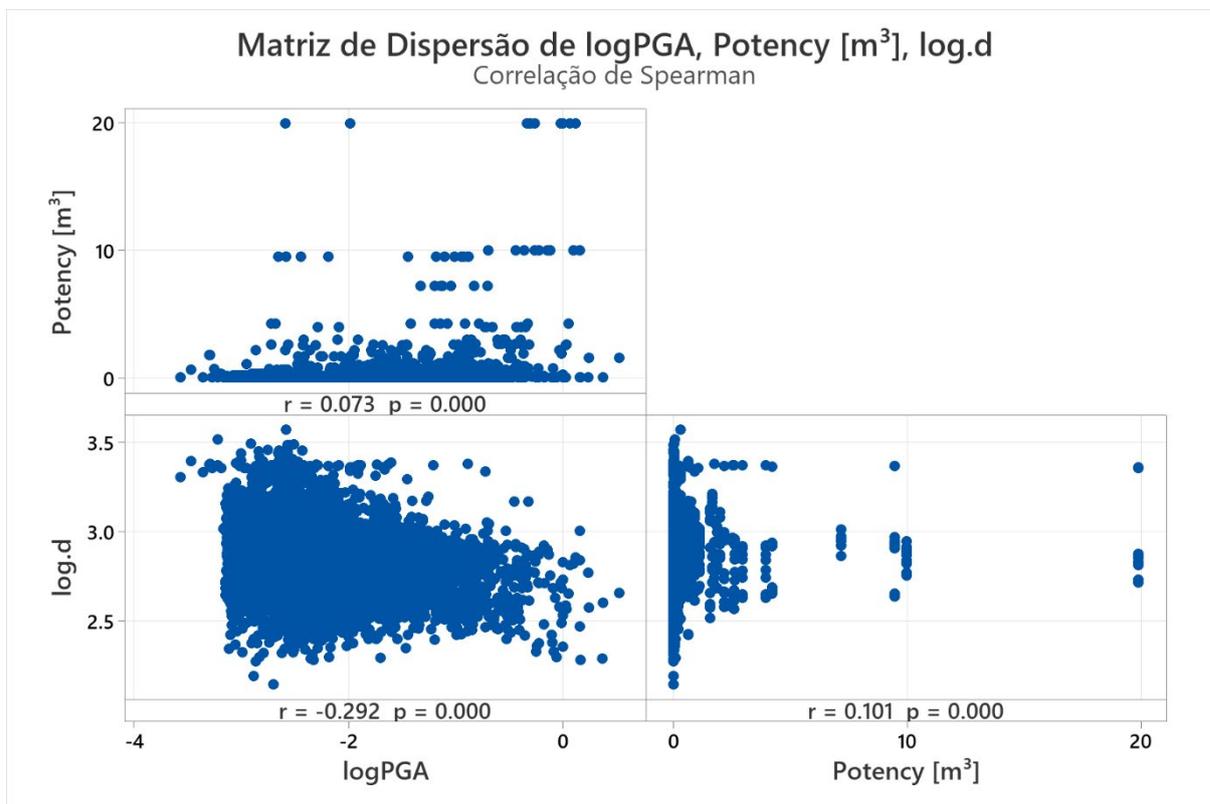
Fonte: Elaborado pela autora.

Observa-se, portanto, que os pressupostos não são francamente respeitados, de forma que não é possível aplicar teste de hipóteses para validar a significatividade das variáveis. Não se trata de resíduos normalmente distribuídos (de acordo com o gráfico de probabilidade normal), tampouco de variância homocedástica (de acordo com os resíduos padronizados versus ajustados).

#### 4.6. REGRESSÃO LINEAR MÚLTIPLA CONFORME LITERATURA

Em virtude das formas comumente adotadas nos modelos de movimentação de terreno (GMM), será adotado o modelo de logaritmo de duas variáveis – PGA e distância, ao passo que potência sísmica permanece em sua forma original (não logarítmica). A matriz de correlação está disponível na Figura 4.25.

Figura 4.25 – Matriz de correlação, entre as log(PGA), log(d) e potência.



### Correlações de Spearman pareadas

Amostra 1	Amostra 2	N	Correlação	IC de 95% para $\rho$	Valor-p
Potency [m <sup>3</sup> ]	logPGA	18447	0.073	(0.059, 0.088)	0.000
log.d	logPGA	18447	-0.292	(-0.305, -0.278)	0.000
log.d	Potency [m <sup>3</sup> ]	18447	0.101	(0.086, 0.115)	0.000

Fonte: Elaborado pela autora.

Observa-se, portanto, que as distribuições se tornaram menos assimétricas, se aproximando da normalidade. No tocante à matriz de correlação entre as três variáveis, os p-valores indicam que não se pode descartar a hipótese de correlação, com p-valores muito próximos a zero.

#### 4.6.1. Avaliação de multicolinearidade

Como supracitado, há correlação entre as variáveis independentes (distância e potência sísmica). Nesse sentido, há chance de multicolinearidade, ao se considerar a amostra de dados empregada.

#### 4.6.2. Resultado da regressão linear múltipla com transformada

O último modelo múltiplo empregado, considerando-se as variáveis PGA e distância transformadas e a variável potência em sua forma original, é linear. Lembra-se que os pressupostos desse modelo são:

- Normalidade dos resíduos;
- Homocedasticidade da variância; e
- Relação de natureza linear entre as variáveis de interesse.

Ou seja, os dados já não cumprem os pressupostos de antemão, haja vista que já se sabe que os dados não apresentam relação de natureza linear. Ainda assim, buscou-se avaliar a regressão linear a fim de se ter sensibilidade de seu coeficiente de determinação preditivo ( $R^2$  pred), que indica quanto da variabilidade da variável resposta pode ser explicada pelo modelo em xequé, que é o comumente aplicado na Sismologia.

Dessa forma, o modelo de regressão linear múltiplo obtido está exposto na Figura 4.26.

Figura 4.26 – Modelo de regressão linear simples para log do PGA predito por log da distância e potência.

#### Equação de Regressão

$$\log\text{PGA} = 0.5478 - 1.0521 \log.d + 0.19411 \text{ Potency [m}^3\text{]}$$

#### Coeficientes

Termo	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	0.5478	0.0718	7.63	0.000	
log.d	-1.0521	0.0252	-41.69	0.000	1.00
Potency [m <sup>3</sup> ]	0.19411	0.00563	34.45	0.000	1.00

#### Sumário do Modelo

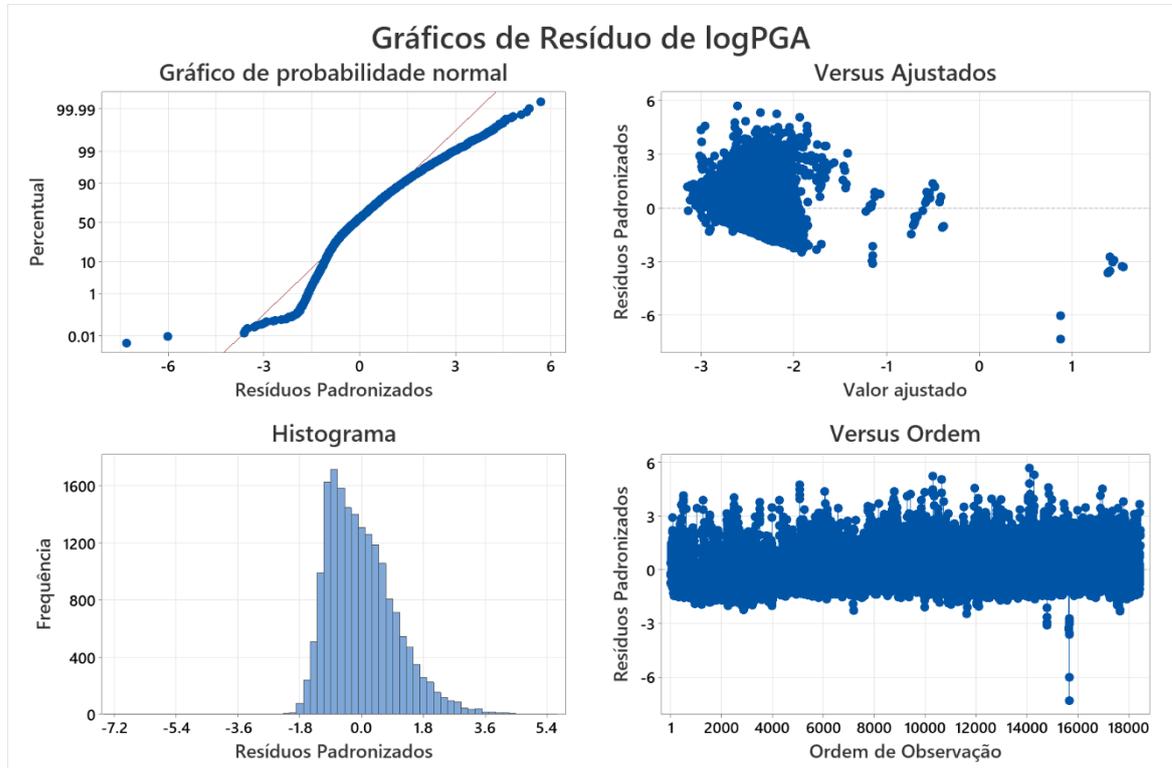
S	R2	R2(aj)	R2(pred)
0.487930	13.09%	13.08%	12.97%

Fonte: Elaborado pela autora.

O coeficiente de determinação preditivo ( $R^2$  pred) obtido foi de 12,97%, menor do que o modelo de regressão linear múltipla considerando-se as variáveis sem transformação ou com todas elas transformadas.

A análise de resíduos é apresentada na Figura 4.27, a seguir.

Figura 4.27 – Análise dos resíduos padronizados para o modelo de regressão linear simples para log de PGA a partir de log da potência sísmica e log da distância.



Fonte: Elaborado pela autora.

Observa-se, portanto, que os pressupostos não são francamente respeitados, de forma que não é possível aplicar teste de hipóteses para validar a significatividade das variáveis, sobretudo em virtude da não normalidade dos resíduos. Não obstante, essa regressão é a que apresenta resultados mais próximos dos pressupostos, com menores valores de assimetria e curtose do histograma e menor enviesamento da variância dos resíduos.

## 5. CONCLUSÃO

O estudo da sismologia tem aplicação em diversas indústrias e um de seus produtos é o modelo de movimentação do terreno (GMM). Ele permite estimar a vibração oriunda de determinado evento sísmico a partir da distância do ponto de interesse à fonte e sua potência sísmica.

É comumente empregada nessa ciência a formulação linear, considerando-se as variáveis originais e/ou com transformações logarítmicas. Assim, optou-se por avaliar a significatividade estatística dessas aplicações em uma amostra de dados de monitoramento microssísmico de uma cava no interior de São Paulo, Brasil. A partir de um período de um ano completo, empregou-se um banco composto de 18.447 trincas de dados – PGA, potência e distância da fonte sísmica.

A partir dos resultados, observou-se que regressões lineares simples explicaram, no melhor cenário, pouco menos de 85% da variabilidade da variável resposta, aceleração de pico do terreno. O modelo de regressão linear múltipla apresentou coeficiente de determinação ligeiramente maior do que os modelos simples, mas ainda na ordem de 15%. Além disso, os pressupostos para verificação da significatividade das variáveis a partir de testes de hipóteses não foram respeitados. Dessa forma, os modelos tiveram pobre caráter preditivo e não é possível validar a regressão.

Com a transformação logarítmica das três variáveis envolvidas, foi novamente construído um modelo linear múltiplo. O resultado do coeficiente de determinação preditivo foi ligeiramente inferior ao que o modelo múltiplo que considerava as variáveis sem transformação. Não obstante, os pressupostos estatísticos também não foram respeitados nesse caso e, assim, não se pode afirmar ser uma regressão válida, além de baixo caráter preditivo.

Por fim, buscou-se elaborar uma regressão múltipla com as variáveis PGA e distância em transformação logarítmica, ao passo que distância foi empregada em sua forma original. De toda forma, os resultados foram similares – baixo coeficiente de

determinação preditivo (12,97%) e falha ao respeitar os pressupostos estatísticos de uma regressão.

A Tabela 1, a seguir, compila os coeficientes de determinação preditivos ( $R^2$  pred) de todos os modelos construídos.

Tabela 1 – Compilação dos coeficientes de determinação preditivos dos modelos construídos

Modelo	$R^2$ pred
Regressão simples $PGA = f(d)$	0,81%
Regressão simples $PGA = f(P)$	14,09%
Regressão múltipla $PGA = f(d, P)$	15,54%
Regressão múltipla $\log(PGA) = f[\log(d), \log(P)]$	15,01%
Regressão múltipla $\log(PGA) = f[\log(d), P]$	12,97%

Fonte: Elaborado pela autora.

Apesar de baixos, os valores de coeficiente de determinação são próximos ao esperado. Dentre os motivos para um valor relativamente baixo, destacam-se:

- Uma base de dados sem distinção da fonte sísmica

Sabe-se que deve haver um modelo para cada tipo de fonte sísmica, em virtude do diferente tipo de onda sísmica que ela emana, implicando em diferentes composições de frequência e, conseqüentemente, diferentes regras de atenuação;

- Heterogeneidade da geologia ao redor da cava

Cada rocha que compõe a geologia local, bem como os efeitos geoestruturais, são descartados na avaliação do modelo de movimentação do terreno (GMM), ao se considerar que a equação seja a mesma em todos os sentidos.

- Falhas no sistema de monitoramento microssísmico da cava

Houver, ao longo do monitoramento, falhas na manutenção dos equipamentos e isso certamente impactou na qualidade do banco de dados.

- Limitações na abordagem estatísticas

Como apontado ao longo do estudo, nenhum dos modelos atendeu, efetivamente, as premissas necessárias para a validade do modelo de regressão. Nesse sentido, deveria ser aplicada uma abordagem não linear, a fim de se respeitar premissas de validade do modelo. Não obstante, o objetivo do presente trabalho foi avaliar os modelos consolidados na literatura sismológica, sendo ele sobretudo o modelo linear considerando-se variáveis transformadas em logaritmo.

De toda forma, com base no presente estudo, não se pode afirmar que as formas de modelos de movimentação do terreno com base em regressões lineares, seja com variáveis originais ou em formato logarítmico, atendem francamente as premissas estatísticas para uma regressão. Ainda que com um banco de dados limitado e de apenas uma região, pode-se afirmar que, para esse caso, seria necessária uma abordagem mais elaborada, o que deve ser objeto de outros trabalhos.

Como sugestão de trabalhos futuros, em virtude do desconhecimento da distribuição dos dados, indica-se a utilização do método de bootstrap não-paramétrico para teste da significância das variáveis independentes do modelo.

## REFERÊNCIAS

- ALARCÓN, J. E. 2003. **Relaciones de atenuación a partir de espectros de respuesta para Colombia**. In: Proceedings of the Second Colombian Conference on Earthquake Engineering. In Spanish.
- AMBRASEYS, N. 1975a. **Ground motions in the near field of small-magnitude earthquakes**. Proceedings of the Commission on the Safety of Nuclear Installations, Organisation of Economic Cooperation in Europe, vol. 1.
- AMBRASEYS, N. N. 1975b. **Trends in engineering seismology in Europe**. Proceedings of Fifth European Conference on Earthquake Engineering, vol. 3.
- AMBRASEYS, N. N. 1978b. **Preliminary analysis of European strong-motion data 1965–1978**. Bulletin of the European Association of Earthquake Engineering, 4.
- BATES, R. L.; JACKSON, J. A. **Dictionary of Geological Terms**. 5. ed. [s.l.] American Geological Institute, 1984.
- COSTA, G., SUHADOLC, P., & PANZA, G. F. 1998. **The Friuli (NE Italy) Accelerometric Network: Analysis of low-magnitude high-quality digital accelerometric data for seismological and engineering applications**. In: Proceedings of the Sixth U.S. National Conference on Earthquake Engineering. Oakland, USA: Earthquake Engineering Research Institute. Seattle, USA. 31 May–4 June.
- DENHAM, D., & SMALL, G. R. 1971. **Strong motion data centre**: Bureau of mineral resources, Canada. Bulletin of the New Zealand Society for Earthquake Engineering, 4(1).
- DENHAM, D., SMALL, G. R., & Everingham, I.B. 1973. **Some strong-motion results from Papua New Guinea 1967–1972**. Proceedings of Fifth World Conference on Earthquake Engineering, vol. 2.
- DOUGLAS, J. 2022. **GMPE Compendium**. Disponível em: < <http://www.gmpe.org.uk/gmpereport2014.html> >. Acesso em: 31 de outubro de 2022.
- FACCIOLI, E. 1979. **Engineering seismic risk analysis of the Friuli region**. Bollettino di Geofisica Teorica ed Applicata, XXI (83).
- GHODRATI AMIRI, G., KHORASANI, M., MIRZA HESSABI, M., & RAZAVIAN AMREI, S. A. 2010. **Ground-motion prediction equations of spectral ordinates and arias intensity for Iran**. Journal of Earthquake Engineering, 14(1).

GRECHKA, V.; HEIGL, W. M. **Microseismic monitoring**. Tulsa/OK - USA: Society of Exploration Geophysics, 2017.

KRAMER, S. L. **Geotechnical earthquake engineering**. [s.l.] Prentice Hall, 1996.

MENDECKI, A. J. **Mine seismology reference book - seismic hazard**. [s.l.] Institute of Mining Seismology, 2016.

MENDECKI, A. J. **Mine seismology: Glossary of selected terms**. Proceeding of the 8th Rockburst and Seismicity in Mines Symposium. Anais...Russia: 2013.

MENDECKI, A. J. **Seismic monitoring in mines**. Londres - Inglaterra: Chapman & Hall, 1997.

MLF GEOMECÂNICA. **Projeto Detalhado - Estruturas da Vale Fertilizantes**. Nova Lima/MG - Brasil: Fundação Coopetec, 5 dez. 2017.

ÓLAFSSON, S., & SIGBJÖRNSSON, R. 1999. **A theoretical attenuation model for earthquake-induced ground motion**. Journal of Earthquake Engineering, 3(3).

PEERC, 2011. **Ground motion prediction equations – 1964-2010**. PACIFIC EARTHQUAKE ENGINEERING RESEARCH CENTER. Disponível em: <[https://apps.peer.berkeley.edu/publications/peer\\_reports/reports\\_2011/webPEER-2011-102-Douglas.pdf](https://apps.peer.berkeley.edu/publications/peer_reports/reports_2011/webPEER-2011-102-Douglas.pdf)>. Acesso em: 31 de outubro de 2022.

PRESS, F. et al. **Para entender a Terra**. [s.l.] Bookman, 2006. v. 4

SEISSOL. SeisSol Home, 15 abr. 2014. Disponível em: <<https://seissol.geophysik.uni-muenchen.de/seissol.html>>. Acesso em: 13 maio. 2020.

SHREARER, P. M. **Introduction to seismology**. Segunda edição ed. New York/NY - USA: Cambridge University Press, 2009.

STEIN, S.; WYSESSION, M. An **introduction to seismology**, earthquakes, and Earth structure. 1. ed. [s.l.] Blackwell Publishing Ltd., 2003.