

# **VALIDAÇÃO DE MODELOS LINEARES**

Uma análise residual

**VINICIUS GOUVEIA DE MIRANDA**

**DOUTORAMENTO EM ENGENHARIA DE MINAS E GEO-AMBIENTE**

Planeamento de Experimentação

PDEMGEO14

Docentes

Antônio Manuel Antunes Fiúza

Joaquim Eduardo Sousa Góis

JULHO DE 2016



## **RESUMO**

O trabalho que segue tem por objetivo estabelecer e definir as principais características de um modelo matemático, com especial ênfase nos modelos lineares. Uma vez definido o modelo, esse necessitará passar por um processo de detecção de “*outliers*” (pontos atípicos) utilizando para isso uma técnica baseada em programação linear, econometria e estatística. Eliminando-se esses pontos do domínio principal será necessário fazer a análise dos resíduos (erro cometido pelo modelo) para comprovar as características intrínsecas a um modelo, como por exemplo, sua natureza normal (distribuição estatística) com esperança nula e variância constante. Uma vez validado o modelo poderemos comparar o modelo tratado pelos métodos propostos com os modelos tradicionais.

**PALAVRAS-CHAVE:** modelo, econometria, programação linear, estatística, resíduos.



# ÍNDICE GERAL

RESUMO .....	III
<b>1. INTRODUÇÃO .....</b>	<b>1</b>
<b>2. MODELOS .....</b>	<b>3</b>
2.1. DEFINIÇÃO .....	3
2.2. FORMA GERAL DE UM MODELO MATEMÁTICO .....	4
<b>3. REGRESSÃO LINEAR.....</b>	<b>8</b>
3.1. ORIGEM HISTÓRICA .....	8
3.2. DEFINIÇÃO .....	9
3.3. SUPOSIÇÕES ASSUMIDAS PARA UM MODELO DE REGRESSÃO LINEAR. ....	17
<b>4. OUTLIERS .....</b>	<b>19</b>
4.1. DEFINIÇÃO INTUITIVA .....	19
4.2. IDENTIFICAÇÃO VISUAL .....	19
4.3. DETEÇÃO DE ATÍPICOS (“OUTLIERS”).....	20
<b>5. VALIDAÇÃO DE MODELOS.....</b>	<b>27</b>
<b>6. CONCLUSÃO E TRABALHOS FUTUROS.....</b>	<b>33</b>
REFERÊNCIAS BIBLIOGRÁFICAS .....	34



## ÍNDICE DE FIGURAS

Figura 1 – Processo de tradução: Realidade para modelo (Goldbarg & L. Luna, 2000) .....	3
Figura 2 - Pontos que representam a topografia de uma frente de bancada .....	4
Figura 3 - Exemplo com dois hiperplanos .....	5
Figura 4 - Hiperplano ajustado à nuvem de pontos .....	5
Figura 5 - Exemplo do resíduo .....	6
Figura 6 - Dados reais e modelo (transladado). .....	7
Figura 7 - Distribuição hipotética das estaturas dos filhos (imagem do artigo original de Karl Pearson).....	8
Figura 8 - Exemplo da soma do quadrado dos resíduos .....	9
Figura 9 - Exemplo de regressão multidimensional.....	11
Figura 10 - Fluxograma da Metodologia para a Estimação da Velocidade de Vibração de Pico das Partículas .....	12
Figura 11 - Gráfico Log(PPV) x Log (SD).....	13
Figura 12 - Regressão por mínimos quadrados .....	14
Figura 13 – Exemplo da soma dos valores absolutos dos resíduos.....	14
Figura 14 - Janela do Excel MS Solver .....	15
Figura 15 - Comparação de métodos de regressão .....	16
Figura 16 - Previsão de vibração como os dois métodos e sua validação dentro da normativa brasileira.....	16
Figura 17 - Distribuição condicional dos resíduos (Gujarati) .....	17
Figura 18 - Comparação entre homoscedasticidade (esq.) e heteroscedasticidade (dir.) (Gujarati).....	17
Figura 19 - Possível ponto atípico.....	19
Figura 20 - Identificação visual de <i>outliers</i> no histograma e no variograma (Czaplicki, 2014) .....	20
Figura 21 - Modelo por mínimos quadrados influenciado pelo <i>outlier</i> e modelo de mínimos resíduos inalterado .....	20
Figura 22 - Planos transladados encontrados por técnicas diferentes de regressão .....	21
Figura 23 - Projeção de ponto no hiperplano .....	22
Figura 24 - Posicionamento de equipamento de <i>perfilometria</i> .....	23
Figura 25 - Falha na tomada de pontos da crista de uma frente livre de bancada. ....	23
Figura 26 - Pontos atípicos resultados de erro de medição .....	24
Figura 27 - Hiperplanos de ajuste: mínimos resíduos absolutos (esq.) e mínimos quadrados (dir.).....	24
Figura 28 - Projeção dos pontos no hiperplano encontrado. ....	25
Figura 29 - Modelo topográfico após tratamento para identificar e eliminar <i>outliers</i> . ....	25

Figura 30 - Foto de uma frente livre e seu modelo tridimensional gerado através de técnicas de fotogrametria.....	26
Figura 31 - Terreno com <i>outliers</i> gerado pela presença de plantas (esq.) e após a eliminação dos <i>outliers</i> (dir.) .....	26
Figura 32 - Exemplo de arquivo gerado por <i>perfilometro</i> 3D antes de qualquer tratamento. 28	
Figura 33 - Tela do Software O-Pitblast, após interpretar os dados do <i>perfilometro</i> 3D .....	28
Figura 34 - Estatísticos descritivos.....	29
Figura 35 - Prova de normalidade para os dados com atípicos.....	29
Figura 36 - Histograma e gráfico Q-Q para dados com atípicos.....	29
Figura 37 - Provas de normalidade para os dados sem atípicos .....	30
Figura 38 - Histograma e gráfico Q-Q para dados sem atípicos.....	30
Figura 39 - Log (PPV) x Log(SD), regressão por mínimos valores residuais e vetores projetados.....	31
Figura 40 - Prova de normalidade para os vetores residuais do modelo de previsão do PPV .....	31
Figura 41 - Resumo do modelo com contraste de <i>Durbin-Watson</i> .....	32
Figura 42 - Diagrama de dispersão dos resíduos.....	32



# 1

## INTRODUÇÃO

O homem vem, desde sempre, tentando representar os fenômenos naturais que o cerca para entender o mundo em geral (Goldbarg & L. Luna, 2000). Dentro de todos os possíveis tipos de modelos enumeramos algumas aplicações importantes citadas por *Tormos et al.* (2003):

- Controle de estoques;
- Teoria de filas;
- Processos de substituição;
- Teoria de jogos;
- Modelos de otimização em redes;
- Gestão de projetos;
- Simulação;
- Programação matemática;
- Processo combinado.

Podemos destacar os que se definem por relações lineares (cujas premissas serão definidas adiante) como os modelos cruciais do século XX devido ao seu importante papel no desenrolar da segunda grande guerra mundial e mais que isso, Hillier & Lieberman (2010) classificam a resolução de problemas lineares (em especial a programação linear) como um dos mais importantes avanços científicos do século passado (uma leitura fantástica e mais detalhada sobre os eventos dos desenvolvimentos matemáticos na segunda guerra pode-se encontrar no “Livro dos códigos”, Singh – 2001).

Embora os modelos lineares sejam comumente utilizados se faz necessário sua validação para poder estimar sua confiança. Essa validação deverá ser realizada baixo alguns critérios, sendo comum, para modelos em geral, que os dados usados sejam divididos em duas partes: uma para o desenvolvimento do modelo e outra, menor, para a validação do mesmo, o que em alguns ramos da ciência é mais mandatário que em outras. Podemos citar essa prática quase que obrigatória na hora de selecionar os melhores modelos de previsão de séries temporais (Peña, 2010, pág. 306). Uma vez estimados os parâmetros do modelo linear se assumem que os erros cometidos (ou resíduos) não constituem um processo de ruído branco (para mais detalhes ler Peña, 2010) e devem seguir uma distribuição de probabilidade normal centrada em zero com variância constante (Gujarati & Porter, 2010, pág. 97). Outra premissa importante para o desenvolvimento desse trabalho será identificar *outliers* (pontos extremos, em geral resultados de erros de medições) no sistema e eliminar esses pontos antes da elaboração dos modelos, pois bastará um ponto para, dependendo da técnica numérica utilizada, ter resultados totalmente errôneos (Hamming R. , 1971, pág. 248).



# 2

## MODELOS

### 2.1. DEFINIÇÃO

O processo de modelagem passa por tentar reproduzir a realidade da forma mais fiel possível, entendendo como se comporta o mundo real e se obtendo as respostas que se podem esperar de determinadas ações. Intuitivamente utilizamos modelos mesmo que não nos demos conta de isso: quando explicamos algo a uma pessoa através de gráficos, ou quando representamos sólidos através de funções matemáticas estamos aplicando modelos. A mente do homem trabalha com estruturas de substituição que tentam simplificar o pensamento, estruturas essas denominadas modelos (Goldbarg & L. Luna, 2000).

Podemos ainda extrapolar a interpretação moderna feita por *Gujarati* (Gujarati & Porter, 2010, pág. 15) e entender um modelo como o estudo da dependência de uma variável (variável dependente) com respeito a uma ou mais variáveis (variáveis explicativas) com o objetivo de estimar ou prever o valor esperado populacional assumindo algumas condições (essas melhores explicadas pelo autor no capítulo 5).

Afinal um bom modelo exige uma correta tradução contextual, e essa pode se expressar através de uma correta relação entre o fenômeno e seu modelo.

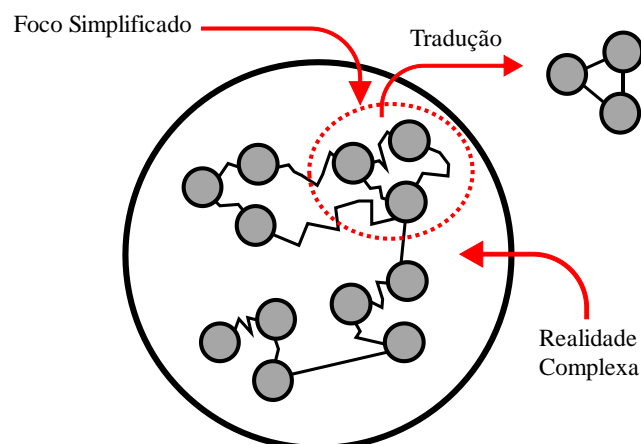


Figura 1 – Processo de tradução: Realidade para modelo (Goldbarg & L. Luna, 2000)

## 2.2. FORMA GERAL DE UM MODELO MATEMÁTICO

Um modelo matemático pode ser considerado como uma arte (parafraseando *Taha* no seu excelente “*Pesquisa Operacional: Uma Visão Geral*”, 2008) e tal qual uma obra de arte, cada modelador, ao desenvolver sua solução, irá criar um conceito diferente, com interpretações diferentes do mesmo problema, mas com um objetivo comum: representar matematicamente um problema abstrato encontrado na natureza.

Um modelo matemático, embora interpretado de maneiras diferentes por diferentes pessoas, terá que compreender principalmente 3 elementos (Tormos Juan & Lova Ruiz, 2003):

1. Variáveis de decisão: o primeiro passo na formulação de um modelo matemático é identificar as variáveis sobre as quais queremos decidir, essas são as variáveis que terão seus valores decididos pelo modelo.
2. Parâmetros de decisão e restrições: os parâmetros podem ser determinísticos ou estocásticos e são os valores conhecidos que se relacionam com as variáveis.
3. Função objetivo: define a medida de efetividade que obtêm o sistema quando se alteram os valores das variáveis de decisão.

Para poder fixar melhor a ideia vamos modelar um problema tipo. Embora uma melhor descrição seja feita *a posteriori*, vamos modelar um conjunto de dados no  $R^3$  (pontos no espaço tridimensional) por um hiperplano <sup>1</sup>.

Inicialmente apresentamos um conjunto de pontos adquiridos com a leitura de um equipamento de leitura topográfica 3D em uma pedreira genérica e que representam uma frente de bancada:



Figura 2 - Pontos que representam a topografia de uma frente de bancada

---

<sup>1</sup> Define-se um hiperplano  $H$  em  $R^n$  como o conjunto de pontos  $(x_1, x_2, \dots, x_n)$  que verificam uma equação linear não degenerada  $a_1x_1 + a_2x_2 + \dots + a_nx_n = b$  (Lipschutz, 1994, pág. 65)

Para encontrar o hiperplano desejado necessitaremos definir quais são nossas variáveis de decisão. Para um plano, temos a equação geral (Boulos, 2005):

$$aX + bY + cZ + d = 0$$

Assim, temos como variáveis de decisão as constantes:  $a$ ,  $b$ ,  $c$  e  $d$ .

Evidentemente que existem infinitas soluções para essa equação, e também uma infinita quantidade de hiperplanos, e aqui exemplificamos dois deles:

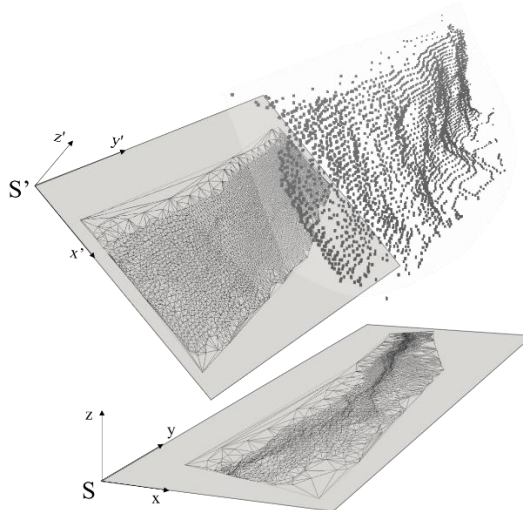


Figura 3 - Exemplo com dois hiperplanos

Para o nosso problema, que é representar nossa nuvem de pontos através do modelo linear “plano” não nos interessa um hiperplano que não se aproxime bem dos pontos, assim que precisaremos encontrar um que seja tão representativo quanto necessitemos.

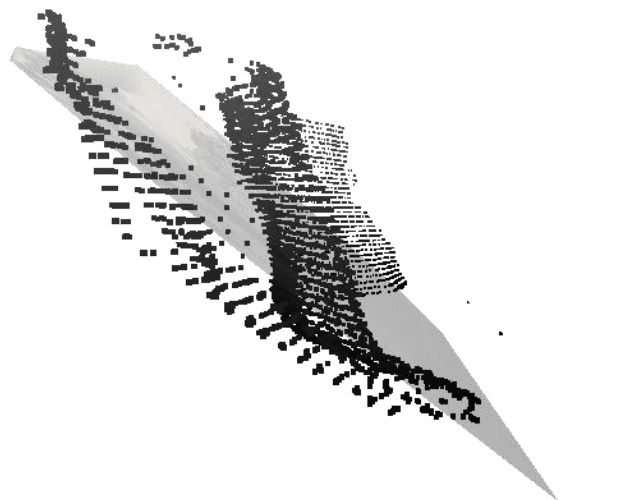


Figura 4 - Hiperplano ajustado à nuvem de pontos

Para alcançar esse objetivo, determinamos o segundo elemento necessário a um modelo, os parâmetros de decisão. Para o nosso caso, sabemos que o modelo comete um erro, pois não coincide (ao menos na grande maioria do seu domínio) com os pontos amostrados, gerando assim um resíduo<sup>2</sup>  $u_i$ .

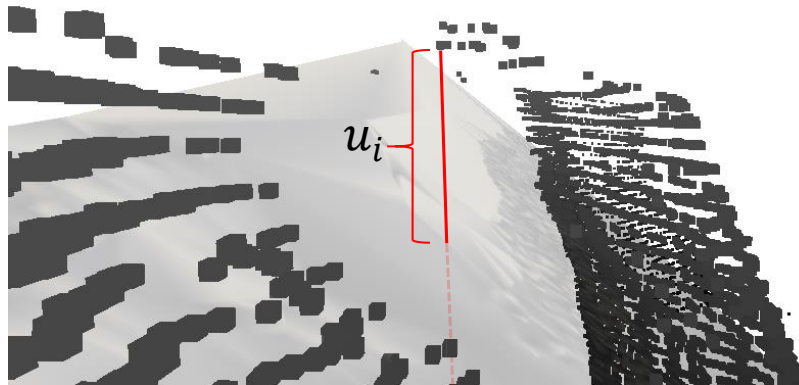


Figura 5 - Exemplo do resíduo

Aqui podemos definir nossos parâmetros:

$$-a/c X_i - b/c Y_i - d/c = AX_i + BY_i + C \quad 1$$

$$AX_i + BY_i + C - Z_i = u_i \quad 2$$

Na equação 1 definimos a relação estocástica entre as variáveis dependentes e a variável independente (Z). Na equação 2 definimos o resíduo como a diferença entre o valor estimado (valor encontrado pelo modelo) e o valor real inferido.

Agora como terceira etapa necessitamos definir a função objetivo, e como queremos que o nosso modelo seja o mais próximo possível da realidade teremos que minimizar os erros cometidos pelo modelo, ou seja, minimizar a soma total dos resíduos. Assim, nossa função objetivo (que deverá ser minimizada) será:

$$\sum_i u_i^2 \quad 3$$

Ou então:

$$\sum_i |u_i| \quad 4^3$$

<sup>2</sup> Perturbação estocástica ou termo de erro estocástico (Gujarati & Porter, 2010, pág. 40).

<sup>3</sup> Embora a equação 4 (abordada mais adiante) minimize a soma total dos resíduos, é raramente utilizada devido a sua complexidade matemática (Hamming, 1971, pág. 249).

Assim, poderíamos definir o modelo matemático de um hiperplano que representa um conjunto de pontos em  $R^3$  como:

Variáveis de decisão:  $A$ ,  $B$  e  $C$ .

$$\text{Min } \sum_i |u_i|$$

S.a.

$$AX_i + BY_i + C - Z_i \leq u_i$$

$$AX_i + BY_i + C - Z_i \geq -u_i$$

$$u_i \geq 0$$

Assim podemos ter um modelo (um espaço vetorial de dimensão inferior) para representar nossos dados reais.

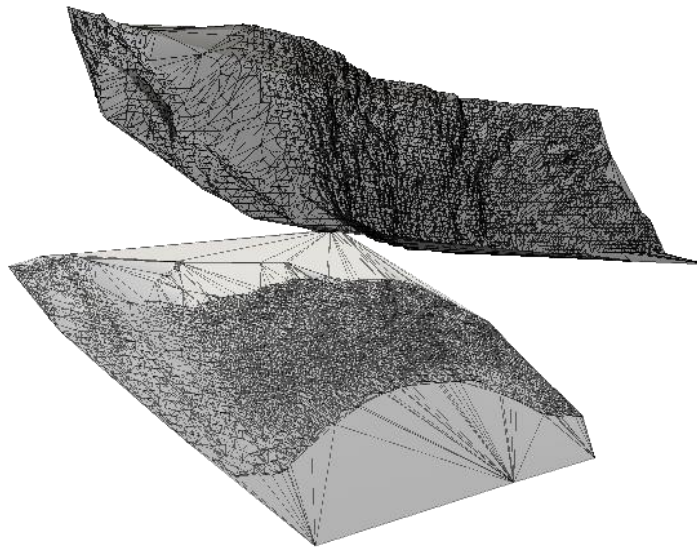


Figura 6 - Dados reais e modelo (transladado).

## 3

## REGRESSÃO LINEAR

## 3.1. ORIGEM HISTÓRICA

Francis Galton, estatístico Inglês do século XIX “batizou” o termo regressão no seu famoso estudo sobre estaturas, onde ele defendia que apesar da tendência de pais altos procriarem filhos altos e pais baixos procriarem filhos baixos, a estatura média dos filhos tendiam a transladar-se, ou “**regressar**” à estatura média da população total (Galton & Hamilton Dickson, 1886, págs. 42-73). Seus estudos foram confirmados pelo seu colega *Karl Pearson*, que investigou estaturas em grupos de famílias (“*Family Record Series*”) e publicou seus resultados em 1903 (Pearson & Lee). Curiosamente, *Galton* se referiu a esse “regresso” à altura média como “regresso à mediocridade” (*Galton em Gujarati*, pág. 15).

A interpretação moderna é diferente, e trata da dependência de uma variável (variável independente) com respeito a uma ou mais variáveis (variáveis explicativas), com o objetivo de estimar o valor médio populacional como bem explica o autor Português *Paulo Jorge Silveira Ferreira* no seu pedagógico “*Princípios de Econometria*” (Editora Rei dos Livros, 2013).

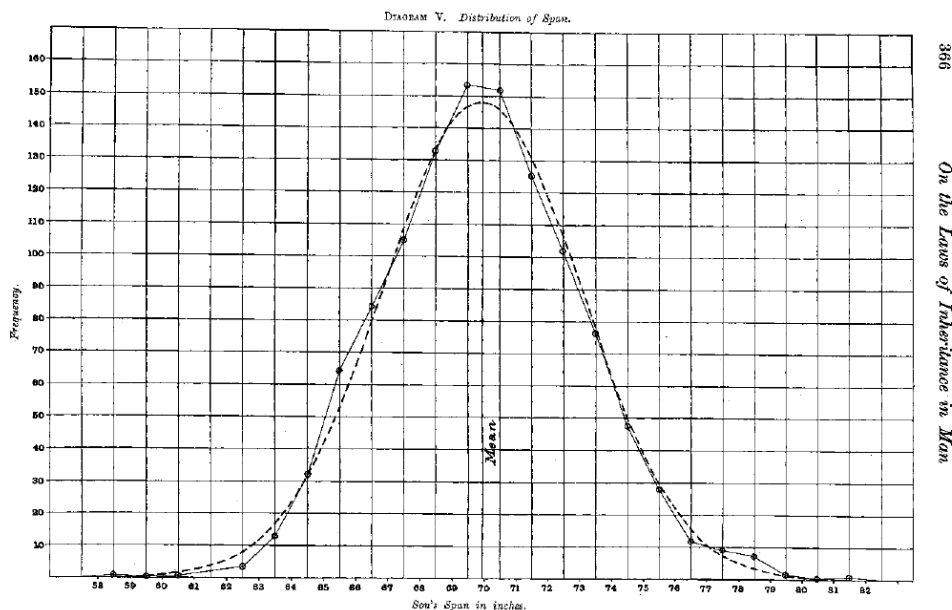


Figura 7 - Distribuição hipotética das estaturas dos filhos (imagem do artigo original de Karl Pearson).



### 3.2. DEFINIÇÃO

Matematicamente definimos um modelo linear com  $n$  variáveis como uma função:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + u_i \quad 5$$

Onde as  $\beta_i$ 's desse modelo são coeficientes lineares, e por linearidade entendemos que três premissas básicas devem ser atendidas (e aqui poderíamos citar todos os autores conhecidos pelo autor desse trabalho: Taha, 2008, pág. 7; Hillier & Lieberman, 2010, pág. 33; Tormos Juan & Lova Ruiz, 2003, pág. 47; Goldberg & Luna, 2000, pág. 47 e o próprio pai (Gass & Assad, 2005, pág. 47) da programação linear moderna, Dantzig no seu clássico, aclamado e único "*Linear Programming and Extensions*", 1963 entre outros):

- Proporcionalidade: aqui definimos que a contribuição individual de cada variável é diretamente proporcional ao valor da variável.
- Aditividade: aqui definimos que a contribuição total das variáveis é igual a soma direta das contribuições individuais de cada variável.
- Certeza: todas as  $\beta_i$ 's são constantes conhecidas e deterministas <sup>4</sup>(aqui não trabalharemos com variáveis estocásticas).

Definido um modelo linear necessitamos encontrar os valores para os parâmetros adequados para ele.

Em ordem de encontrar os parâmetros que melhor se ajustam aos nossos pontos precisaremos aplicar alguma técnica, e a seguir listamos o método dos mínimos quadrados, largamente difundido e usado, e o método dos mínimos resíduos absolutos, que não é explorado pela literatura econométrica/matemática, em parte pela sua dificuldade técnica (Hamming, 1989) mas que será bastante usado pelo autor desse trabalho.

#### 3.2.1. MÉTODO DOS MÍNIMOS QUADRADOS

##### 3.2.1.1. Caso Bidimensional

Como já discutido no apartado 2.2 desse trabalho um modelo explica parte da realidade e uma parte que não é explicada é chamada de resíduo (a partir de agora chamado de " $u$ ").

A ideia por trás do método dos mínimos quadrados é encontrar os parâmetros que minimizam a soma do quadrado dos resíduos, ou seja, nosso objetivo será:

$$\text{Min } \sum_i u_i^2 \quad 6$$

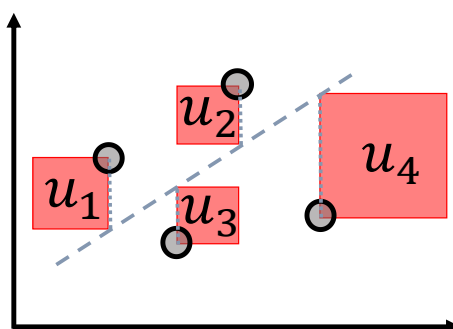


Figura 8 - Exemplo da soma do quadrado dos resíduos

<sup>4</sup> Existem modelos de regressão, chamados **modelos neoclássicos de regressão linear (MNRL)** que tratam as variáveis como estocásticas (Goldberger, 1991, pág. 264)

Observasse que infinitas retas podem ser escolhidas, mas apenas uma fará com que a soma dos quadrados dos resíduos (Figura 8) seja a mínima possível<sup>5</sup>.

Os parâmetros são encontrados da seguinte maneira:

$$y = \beta_0 + \beta_1 X + u \quad 7$$

Onde:

- $\beta_0$ : parâmetro chamado de constante
- $\beta_1$ : coeficiente técnico (Tormos Juan & Lova Ruiz, 2003, pág. 35)
- $u$ : Resíduo

O objetivo é minimizar o quadrado dos resíduos:

$$\min \sum u_i^2 = \min \sum_{i=1}^n (y_i - \beta_0 - \beta_1 X_i)^2 = S(\beta_0, \beta_1) \quad 8$$

A minimização se dá ao derivar a equação 8 com relação a  $\beta_0$  e a  $\beta_1$  e igualando a zero:

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 X_i) = 0 \quad 9$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n X_i (y_i - \beta_0 - \beta_1 X_i) = 0 \quad 10$$

Distribuindo corretamente e dividindo a expressão 9 por  $2n$  chegamos a:

$$\beta_0 = \bar{y} - \beta_1 \bar{X} \quad 11$$

Onde  $\bar{y}$  é a média dos valores da variável dependente, e  $\bar{X}$  é a média da variável independente.

Substituindo a equação 11 na equação 10 chegamos a:

$$-2 \sum_{i=1}^n X_i (y_i - \bar{y} + \beta_1 \bar{X} - \beta_1 X_i) = 0 \quad 12$$

$$\sum_{i=1}^n [X_i (y_i - \bar{y}) + X_i \beta_1 (\bar{X} - X_i)] = 0 \quad 13$$

$$\sum_{i=1}^n [X_i (y_i - \bar{y}) + \beta_1 \sum_{i=1}^n X_i (\bar{X} - X_i)] = 0 \quad 14$$

$$\beta_1 = \frac{\sum_{i=1}^n X_i (y_i - \bar{y})}{\sum_{i=1}^n X_i (X_i - \bar{X})} \quad 15$$

### 3.2.1.2. Caso Multidimensional

Para o caso multidimensional vamos resolver o mesmo problema para:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + u_i \quad 16$$

Para poder realizar isso necessitamos escrever a equação 16 na forma matricial:

<sup>5</sup> Observasse aqui que a técnica **não** (e aqui fazemos especial destaque) nos resulta na reta que melhor aproxima os pontos, pois ela não minimiza a distância dos pontos até a reta (ou modelo), e sim a distância elevada ao quadrado.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ 1 & X_{13} & X_{23} & \dots & X_{k3} \\ 1 & X_{14} & X_{24} & \dots & X_{k4} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ \dots \\ e_n \end{pmatrix} \quad 17$$

Podemos simplificar essa fórmula:

$$y = Xb + e \quad 18$$

Substituindo e por  $y - Xb$ , temos e considerando que queremos minimizar a soma do quadrado dos resíduos:

$$\sum_i e_i^2 = \sum_i (e'e)^2 \quad 19$$

Substituindo e por  $y - Xb$ , temos:

$$\frac{\partial S}{\partial b} = -2X'y + 2X'Xb = 0 \quad 20$$

$$X'Xb = X'y \quad 21$$

$$b = (X'X)^{-1}X'y \quad 22$$

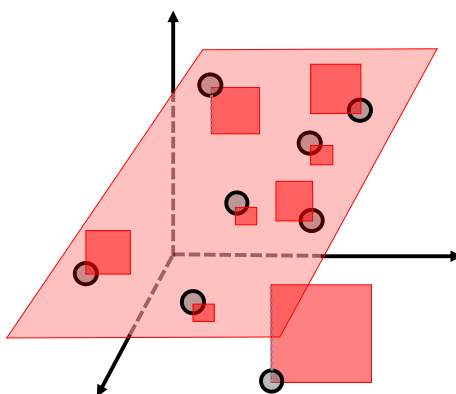


Figura 9 - Exemplo de regressão multidimensional

### 3.2.1.3. Exemplo de aplicação

As minerações, pedreiras e construções produzem certo grau de vibração devido a energia que não se utiliza no rompimento da rocha (Siskind, 2005), e ainda podem originar conflitos indesejados e queixas, devido as muitas ocasiões nas quais os desmontes realizados se produzem em zonas próximas as áreas urbanas (Sanchidrián & Muñiz, 2000). Pelo exposto, fica claro a necessidade de se poder modelar e prever as vibrações e assim poder evitar danos à população civil.

Dowding exemplifica como realizar um estudo de vibrações no terceiro capítulo de “*Blast Vibration Monitoring and Control*” (1985) e, de maneira simplificada, podemos definir que o nível de vibração (a partir daqui chamado de *Peak Particle Velocity*, ou simplesmente PPV) pode ser estimado por:

$$PPV \left( \frac{mm}{s} \right) = KQ^\alpha D^\beta \quad 23$$

Onde:

- K é um regressor que basicamente é influenciado pelo terreno por onde se propaga o evento sísmico.
- Q é a carga máxima (Kg) utilizada no desmonte.
- $\alpha$  é um regressor que corrige a carga.
- D é a distância (m) euclídea entre o ponto de detonação e o local onde se pretende prever o PPV.
- $\beta$  é o regressor que corrige a distância.

A metodologia para um estudo de vibrações foi abordada por Andrade Gomes (2016) em sua dissertação para obtenção do título de Mestre em Engenharia de Minas pela Universidade do Porto em 2016 intitulada “Metodologia para Análise de Vibrações Provocadas por Desmonte com Recurso a Explosivos” (coorientada pelo autor desse trabalho). Ali podemos encontrar um fluxo geral sobre o estudo de vibrações:

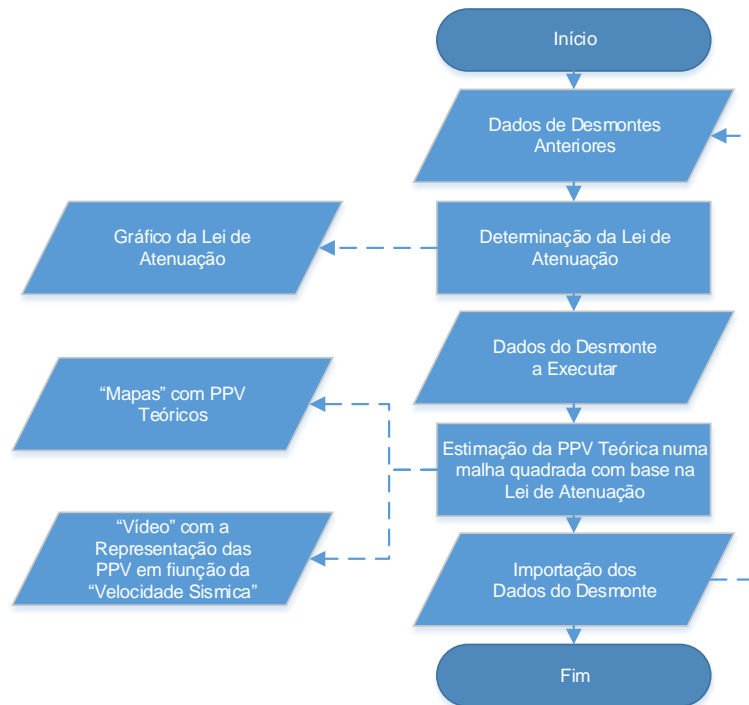


Figura 10 - Fluxograma da Metodologia para a Estimação da Velocidade de Vibração de Pico das Partículas

Como se pode observar a determinação da lei de atenuação é uma etapa crucial, e a Regressão linear é fundamental para sua execução.

Para esse exemplo vamos reproduzir parte dos dados de campo publicados por Andrade apenas para exemplificar, mas para isso faremos inicialmente uma redução de dimensão, e assumiremos que:

$$\beta = -2 * \alpha \quad 24$$

$$SD = D / \sqrt{Q} \quad 25$$

Essa simplificação é usual para eventos sísmicos originados por cargas cilíndricas e é comumente exemplificado na literatura (Siskind, 2005, pág. 21).

Embora a natureza da equação 23 seja não linear (não obedece as restrições de linearidade citadas no apartado 3.2), a equação pode ser linearizada (Mathews & Fink, 2000, pág. 292). Aplicando a simplificação e linearização podemos chegar a:

$$\log_{10} PPV = \log_{10} K + \beta \log_{10} SD \quad 26$$

Que é linear e uma vez que tenhamos um conjunto de dados poderemos regressar os valores de  $K$  e de  $\beta$ .

Usaremos aqui os dados (simplificados) apresentados por Andrade (pág. 44):

PPV (mm/s)	Distância (m)	Carga (Kg)	SD (m/Kg <sup>0.5</sup> )	Log (PPV)	Log (SD)
12.472 mm/s	204 m	84 Kg	22.3	1.096 mm/s	1.348
13.386 mm/s	231 m	187 Kg	16.9	1.127 mm/s	1.228
1.140 mm/s	243 m	70 Kg	29.0	0.057 mm/s	1.463
7.036 mm/s	246 m	75 Kg	28.4	0.847 mm/s	1.453
7.036 mm/s	246 m	145 Kg	20.4	0.847 mm/s	1.310
12.627 mm/s	272 m	159 Kg	21.6	1.101 mm/s	1.334
11.808 mm/s	426 m	143 Kg	35.6	1.072 mm/s	1.552
5.964 mm/s	565 m	310 Kg	32.1	0.776 mm/s	1.506
6.678 mm/s	632 m	198 Kg	44.9	0.825 mm/s	1.653
1.535 mm/s	1486 m	210 Kg	102.5	0.186 mm/s	2.011
1.403 mm/s	1538 m	191 Kg	111.3	0.147 mm/s	2.046
1.092 mm/s	1588 m	149 Kg	130.1	0.038 mm/s	2.114

Tabela 1 - Dados de vibrações medidos em campo

Aqui podemos montar, baseado nos dados, o gráfico que representa o PPV em função do SD (claro, em logaritmos para garantir a natureza linear do modelo):

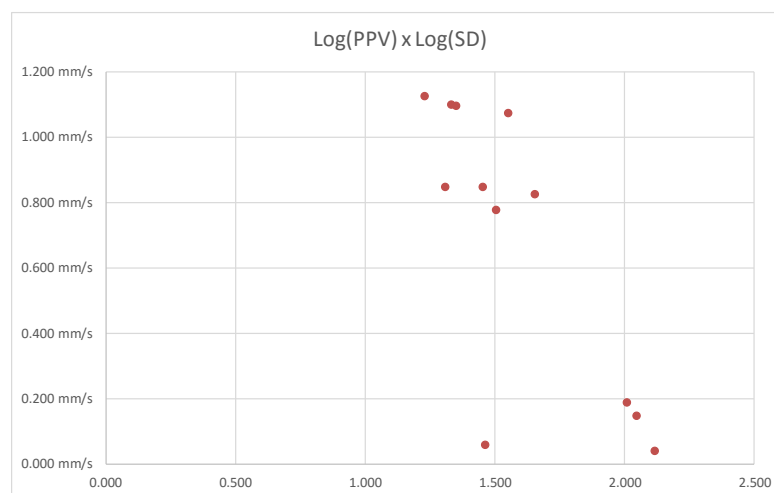


Figura 11 - Gráfico Log(PPV) x Log (SD)

Aplicando as equações desde 9 até a equação 15 podemos encontrar o modelo linear que ajusta esses pontos:

$$PPV = 279.77 SD^{-1.117}$$

27

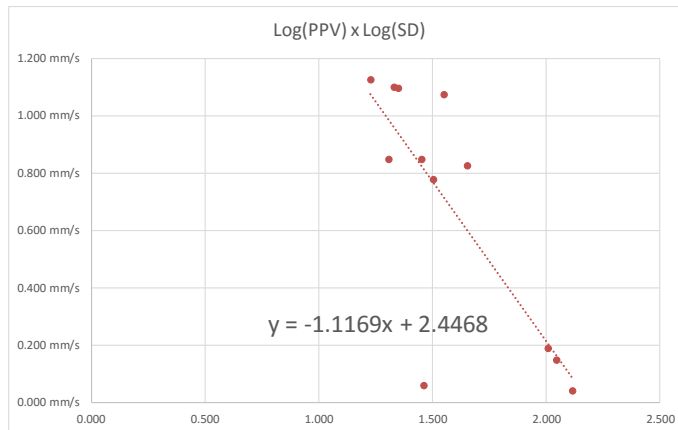


Figura 12 - Regressão por mínimos quadrados

Aqui modelamos nossos dados do espaço bidimensional por uma reta (ou um hiperplano ajustado).

### 3.2.1. MÉTODO DOS MÍNIMOS RESÍDUOS

Essa técnica consiste em minimizar a soma do valor absoluto do erro (resíduo) cometido:

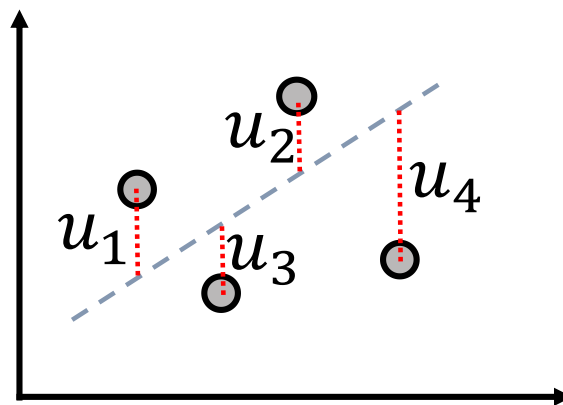


Figura 13 – Exemplo da soma dos valores absolutos dos resíduos.

A resolução desse problema (minimização da função objetivo definida pela equação 4) é de natureza matemática bastante complicada como *Hamming* definiu em 1979, e embora o próprio *Hamming* tente abordar esse problema (*Hamming R.*, 1973, pág. 431) sua solução não é apresentada. Para superar esse problema utilizaremos programação linear.

Uma vez conhecidos as distâncias escalonadas (“*Scaled Distances-SD*”), nosso propósito é minimizar a soma do valor absoluto dos resíduos. Para isso devemos encontrar as variáveis de decisão  $A$ ,  $B$  e  $u_i$  resolvendo o modelo (para o caso bidimensional):

$$\min \sum_i u_i \tag{28}$$

S.a.

$$y_i - \beta_0 - \beta_1 X_i \leq u_i \tag{29}$$

$$y_i - \beta_0 - \beta_1 X_i \geq -u_i \tag{30}$$

$$u_i \geq 0 \tag{31}$$

Para resolver esse problema a quantidade de variáveis utilizadas será fundamental, pois um problema muito grande pode-nos levar à decisão de não buscar a melhor solução e sim uma solução factível e razoavelmente próxima da solução ótima (Hillier & Lieberman, 2010, pág. 563).

Para problemas lineares de até 200 variáveis de decisão é possível utilizar o Microsoft solver (Microsoft, 2016), que apesar de não possuir a robustez que pode possuir o CPLEX (IBM, 2016) nem a potência do pacote estatístico de programação linear do SAS (Institute, 2016) é acessível e eficiente para pequenos problemas.

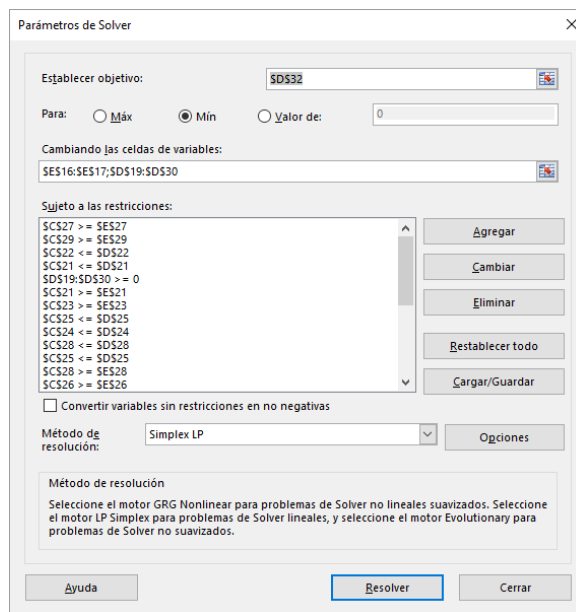


Figura 14 - Janela do Excel MS Solver

Uma vez introduzidos os parâmetros no MS Solver, a solução encontrada difere da solução por mínimos quadrados, tendo como solução:

$$PPV=399.85 \text{ } SD^{-1.2013} \tag{32}$$

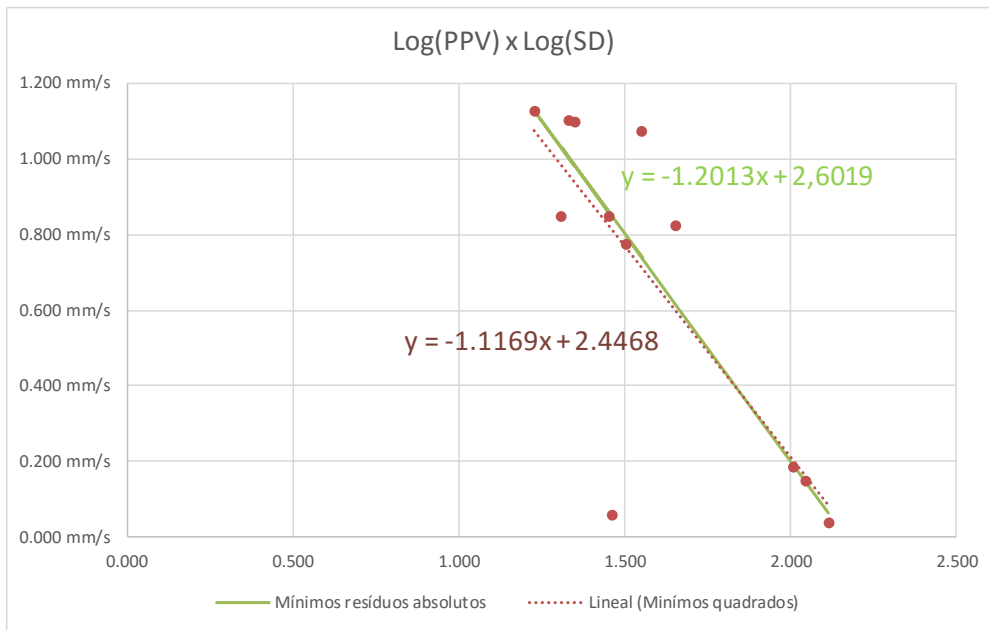


Figura 15 - Comparação de métodos de regressão

Embora essa diferença pareça pequena, elas têm um impacto muito grande e podem fazer a diferença entre ultrapassar ou não uma normativa legal.

Exemplificaremos com um caso. O grau de vibrações (PPV) que se admite legalmente na normativa Brasileira é uma função que depende não apenas de sua intensidade como também da frequência do fenômeno vibratório (ABNT, 2005). Uma frequência de 17 Hz é comum em desmontes tanto de minas de carvão, como em pedreiras e obras civis (Dowding, 1985, pág. 18). Se estipularmos uma carga de 100 Kg e analisarmos seu previsto efeito a uma distância de 100 metros (assumindo a mesma direção com a qual os dados foram obtidos) teríamos uma regra dentro da normativa enquanto a outra não.

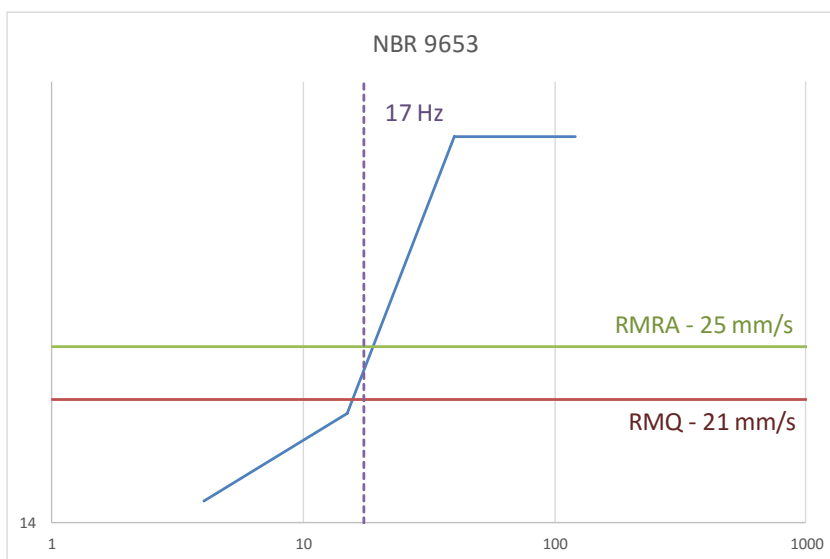


Figura 16 - Previsão de vibração com os dois métodos e sua validação dentro da normativa brasileira.



### 3.3. SUPOSIÇÕES ASSUMIDAS PARA UM MODELO DE REGRESSÃO LINEAR.

O **modelo de Gauss, modelo clássico ou estândar de regressão linear (MCRL)** é o cimento da teoria econométrica (Gujarati & Porter, 2010, pág. 61), e nos fornece 7 suposições que nossos modelos deverão (ou tentarão) enquadrar. Estendemos o entendimento, aplicado ao método dos mínimos quadrados ao método dos mínimos resíduos absolutos.

#### 3.3.1. REGRESSÃO LINEAR

O modelo de regressão será sempre linear **nos parâmetros** (isto é, os parâmetros  $\beta_1, \beta_2, \dots, \beta_n$ ).

#### 3.3.2. VARIÁVEL INDEPENDENTE NÃO RELACIONADAS COM O RESÍDUO

Essa suposição assume que a variável (ou as variáveis) independentes possuem covariância nula com respeito aos resíduos:  $\text{cov}(X_i, u_i) = 0$ .

#### 3.3.3. O VALOR ESPERADO DO RESÍDUO É ZERO

A média dos resíduos que estão ao redor do mesmo ponto (ou pontos) do domínio é zero.

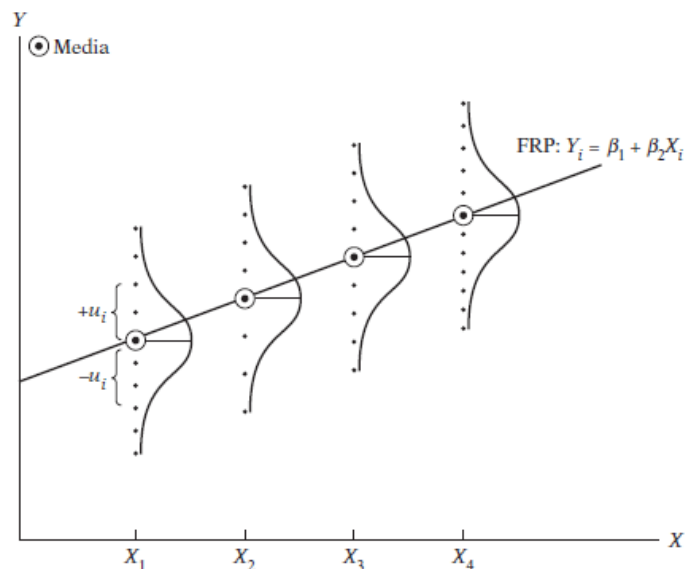


Figura 17 - Distribuição condicional dos resíduos (Gujarati)

#### 3.3.4. HOMOSCEDASTICIDADE OU VARIÂNCIA CONSTANTE

A variância  $\sigma^2$  residual é constante e não se altera ao longo do domínio.

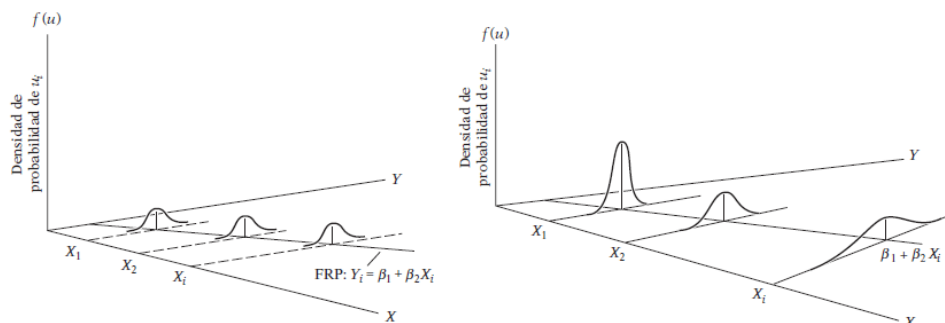


Figura 18 - Comparação entre homoscedasticidade (esq.) e heteroscedasticidade (dir.) (Gujarati)

### 3.3.5. NÃO EXISTE CORRELAÇÃO ENTRE OS RESÍDUOS

Dados dois valores dos resíduo,  $u_i$  e  $u_j$ , por exemplo,  $cov(u_i, u_j)=0$ .

### 3.3.6. NÚMERO DE OBSERVAÇÕES

O número de observações  $n$  deve ser maior que o número de parâmetros a estimar, pois do caso contrário os parâmetros seriam indetermináveis.

### 3.3.7. NATUREZA DA VARIÁVEL (OU VARIÁVEIS) INDEPENDENTE

Não devem existir valores atípicos para a variável independente, quer dizer, valores que se distanciam muito da média amostral.

Não existe consenso sobre a necessidade de cumprir o não essas sete suposições, e não seria raro (na verdade é mais ao contrário disso) encontrar modelos que não cumprem todas as premissas. Para mais detalhes podemos consultar o capítulo 3 de *Novales* (1993).

# 4

## OUTLIERS

### 4.1. DEFINIÇÃO INTUITIVA

Quando um pesquisador experiente tem claro que uma das amostras tomada para sua pesquisa tem um considerado desvio com relação a população (estatística) ele sabe que essa observação tem que ser descartada (Grubbs, 1969), mas quando esse desvio não é evidente se fazem necessários alguns critérios estatísticos para sua detecção. Czaplicki (2014, pág. 58) observa que na toma de uma amostra alguns valores claramente diferem dos outros (da grande maioria), e em matemática ou estatística esses valores são comumente chamados “*outliers*”.

As definições de *outliers* e a matemática por trás de sua identificação são inúmeras, desde as mais simples como a exemplificada por Bruce Simmons (Simmons, 2016) baseada nos quartis, como algumas mais elaboradas que passam por normalização de dados (ainda em acordo com Grubbs, pág. 3).

### 4.2. IDENTIFICAÇÃO VISUAL

Alguns pontos são, como já defendido, visivelmente erros de medições, ou resultados de situações atípicas que podem mascarar os resultados.

Nos dados analisados no capítulo anterior podemos identificar claramente um possível valor atípico:

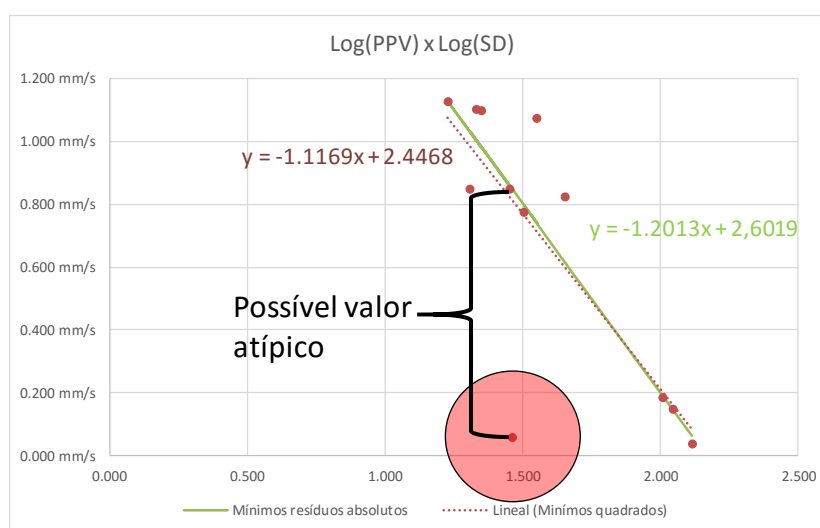


Figura 19 - Possível ponto atípico

É evidente que esses pontos tendem a distorcer nosso modelo, pois tendem a dominar o resultado, fato esse evidenciado no método dos mínimos quadrados (Hamming R. , 1971, pág. 248) (Hamming R. , 1973, pág. 431).

Czaplicki ainda ressalta (pág. 58) a importância do histograma para poder identificar visualmente possíveis pontos atípicos.

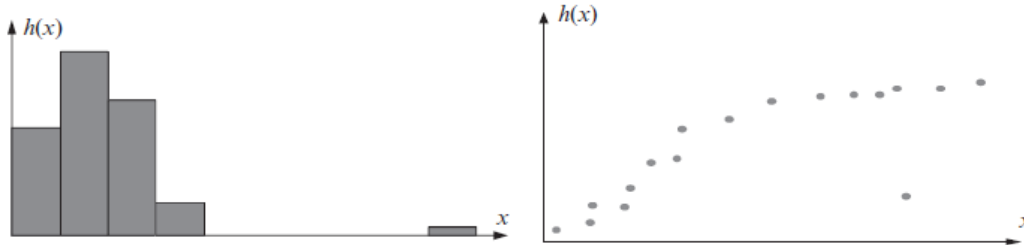


Figura 20 - Identificação visual de outliers no histograma e no variograma (Czaplicki, 2014)

### 4.3. DETEÇÃO DE ATÍPICOS (“OUTLIERS”)

A primeira etapa do processo de detecção dos outliers é determinar um hiperplano que aproxime os pontos do espaço linear em questão. Sabemos que um hiperplano para o espaço bidimensional  $R^2$  é uma reta, e que para o espaço tridimensional  $R^3$  é um plano (Lipschutz, 1994), e podemos também generalizar que para qualquer problema n-dimensional iremos encontrar um hiperplano (n-1)-dimensional.

Uma das tarefas importantes aqui é entender que para tal o método dos mínimos quadrados não é o mais apropriado, pois o mesmo é fortemente influenciado pelos outliers, como podemos observar quando deslocamos o ponto da nuvem original e vemos que o comportamento do hiperplano quando calculado por mínimos quadrados se altera muito, enquanto com mínimos resíduos não.

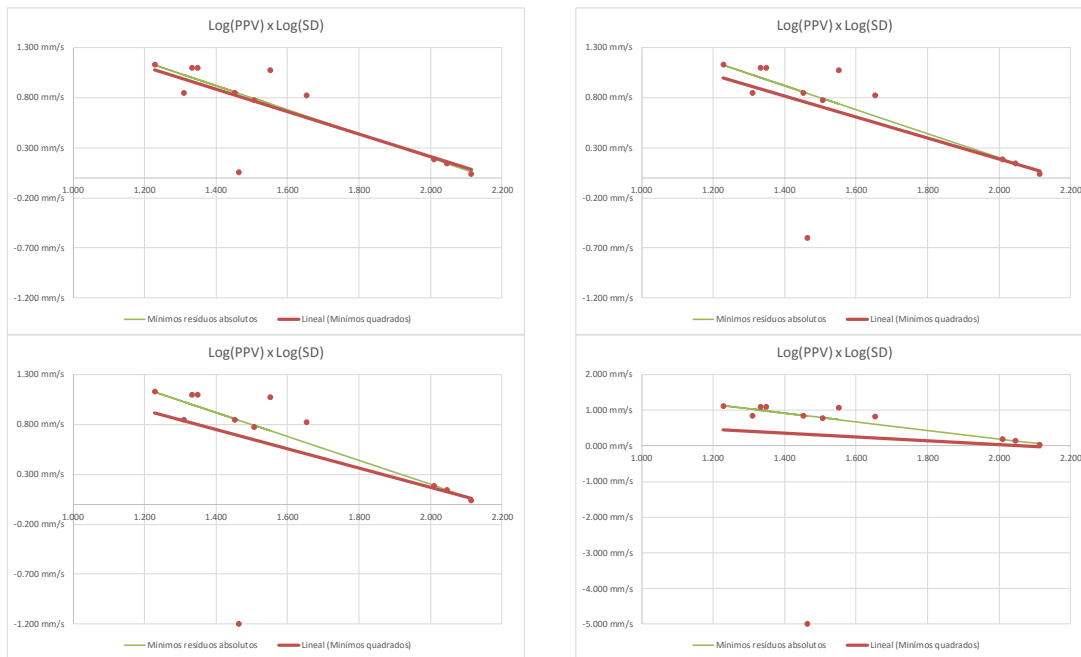


Figura 21 - Modelo por mínimos quadrados influenciado pelo outlier e modelo de mínimos resíduos inalterado

Podemos usar para exemplificar o espaço tridimensional os pontos que formam um terreno. O hiperplano nesse caso será um plano.

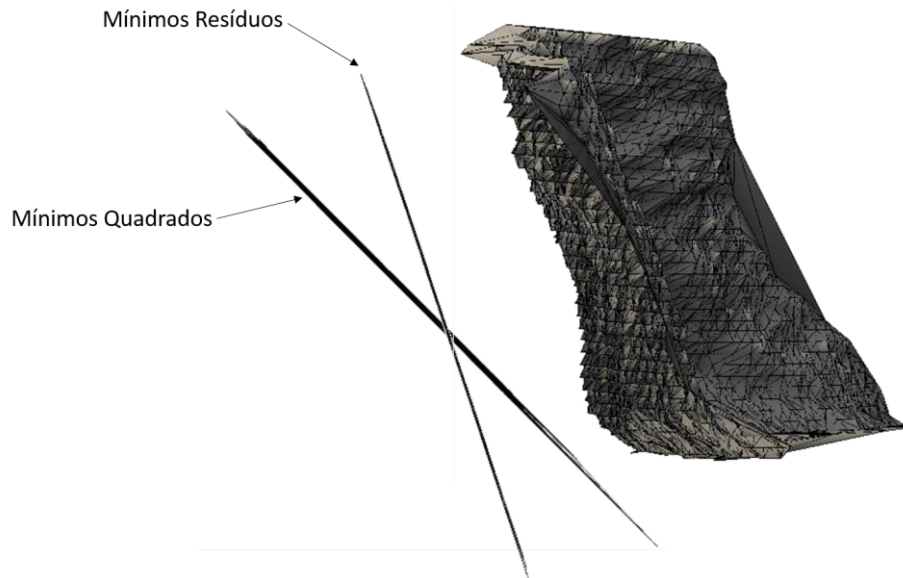


Figura 22 - Planos transladados encontrados por técnicas diferentes de regressão

Encontrado esse hiperplano<sup>6</sup>, deveremos utilizar uma regressão por mínimos resíduos e projetar sobre esse hiperplano os pontos originais.

Para poder, genericamente, fazer essa projeção, deveremos proceder como segue:

Seja P um ponto do espaço  $R^n$   $\langle P_1, P_2, \dots, P_n \rangle$ , e esse espaço construído sobre uma base E  $\langle \hat{e}_1, \hat{e}_2, \dots, \hat{e}_n \rangle$ , temos a equação do plano encontrado  $\pi$ :  $\beta_0 + \beta_1 \hat{e}_1 + \beta_2 \hat{e}_2 + \dots + \beta_n \hat{e}_n = 0$ .

Temos o vetor  $\vec{N} = \langle \beta_1, \beta_2, \dots, \beta_n \rangle$  um vetor ortogonal ao hiperplano encontrado (o produto vetorial entre um vetor diretor do hiperplano e  $\vec{N}$  tem como resultado um outro vetor diretor do hiperplano  $\pi$ ).

Temos também n-1 vetores diretores de  $\pi$ :  $u^1, u^2, \dots, u^{n-1}$ .

Temos um ponto qualquer pertencente a  $\pi$ :  $\langle \pi_1, \pi_2, \dots, \pi_n \rangle$ .

Agora montamos o seguinte sistema de equações:

$$P_1 + \beta_1 \lambda^1 = \pi_1 + u_1^1 \lambda^2 + u_1^2 \lambda^3 + \dots + u_1^{n-1} \lambda^n$$

$$P_2 + \beta_2 \lambda^1 = \pi_2 + u_2^1 \lambda^2 + u_2^2 \lambda^3 + \dots + u_2^{n-1} \lambda^n$$

...

$$P_n + \beta_n \lambda^1 = \pi_n + u_n^1 \lambda^2 + u_n^2 \lambda^3 + \dots + u_n^{n-1} \lambda^n$$

33

Agora montamos o sistema matricialmente para poder melhor visualizar e resolver o sistema de equações.

<sup>6</sup> Insistimos em referirmos ao modelo como hiperplano e não como plano para não gerar confusão entre o espaço de dimensão inferior e o espaço bidimensional.

$$\begin{pmatrix} \beta_1 & -u_1^1 & -u_1^2 & \cdots & -u_1^{n-1} \\ \beta_2 & -u_2^1 & -u_2^2 & \cdots & -u_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_n & -u_n^1 & -u_n^2 & \cdots & -u_n^{n-1} \end{pmatrix} \begin{pmatrix} \lambda^1 \\ \lambda^2 \\ \vdots \\ \lambda^n \end{pmatrix} = \begin{pmatrix} \pi_1 - P_1 \\ \pi_2 - P_2 \\ \vdots \\ \pi_n - P_n \end{pmatrix} \quad 34$$

Podemos simplificar a equação acima como:  $Xa=Y$ .

Então a solução do sistema acima será:

$$a = X^{-1}Y \quad 35$$

E substituindo o primeiro elemento do vetor “a” (35) nos elementos do lado esquerdo do sistema de equações 33 encontramos a projeção do ponto no hiperplano.

Uma vez que projetamos os pontos no hiperplano, deveremos encontrar a longitude do vetor (distância euclídea) entre o ponto projetado e o ponto original.

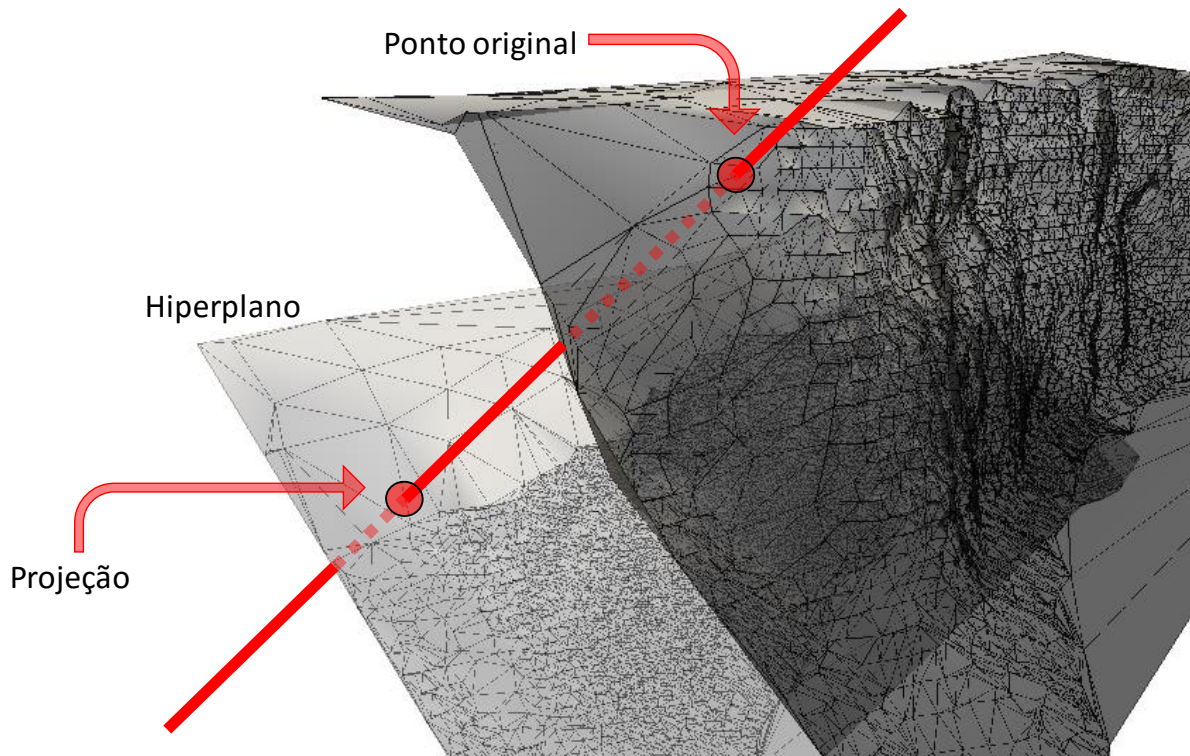


Figura 23 - Projeção de ponto no hiperplano

Assumiremos aqui que (e discutiremos com mais detalhe no seguinte capítulo desse trabalho) a normalidade <sup>7</sup>das longitudes dos vetores formados pelos pontos e suas projeções.

<sup>7</sup> Aqui encontramos uma discordância sobre a natureza normal dos resíduos entre *Hamming* (1973, pág. 432) e os econométricos (Gujarati & Porter, 2010, pág. 97), mas assumimos (pela extensão dos trabalhos) a visão econométrica como mais adequada (Novales Cinca, 1993), (Pearson & Lee, 1903) entre outros, que defendem a normalidade dos dados em questão.

Aqui devemos assumir que os dados são normais truncados (estritamente positivos), uma vez que tratamos com longitudes de vetores.

Aqui devemos calcular a média  $\mu$  e a variância  $\sigma^2$  da amostra.

Seguindo a regra dos 3 sigmas<sup>8</sup> (Pukelsheim, 1994), sabemos que os valores, com um nível de significância de  $x\%$  e um  $\alpha = 1 - x\%$ , deve estar compreendido dentro do intervalo:

$$[\mu + \xi_{\alpha/2}^{-1} \sigma ; \mu - \xi_{\alpha/2}^{-1} \sigma], \text{ onde } \xi_{\alpha/2}^{-1} \text{ é a distribuição normal inversa.} \quad 36$$

Os valores que estiverem fora desse intervalo serão considerados *outliers* e serão excluídos do modelo.

Aqui temos um exemplo para melhor entendimento: Considere um Engenheiro de minas que precisa avaliar as condições da bancada antes de uma detonação usando um equipamento de *perfilometria* (Bhandari, 1997, pág. 169). O engenheiro posicionará o laser em frente a bancada e começará a tomar medidas.

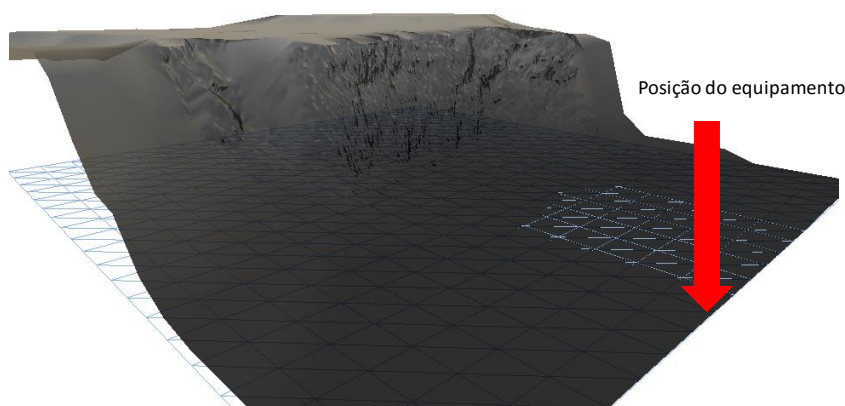


Figura 24 - Posicionamento de equipamento de *perfilometria*

Nessa situação é comum que o engenheiro, ao tomar as pontas da crista, tome uma medição mais alta que a devida e capte um ponto errôneo:

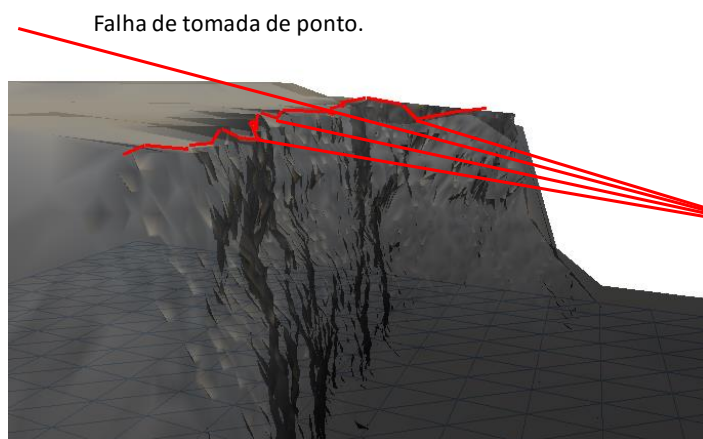


Figura 25 - Falha na tomada de pontos da crista de uma frente livre de bancada.

<sup>8</sup> Embora o modelo se refira a 3 sigmas (99,73% do intervalo compreendido) e exista na literatura específica da engenharia de minas referência aos 3 desvios (Czaplicki, 2014, pág. 67) se faz necessário uma calibração. Para terrenos, o melhor ajuste encontrado pelo autor desse trabalho se encontra com 2.32 desvios (98% do intervalo compreendido):

O resultado desse erro operacional resulta num terreno que não corresponde às necessidades do engenheiro de minas naquele propósito:

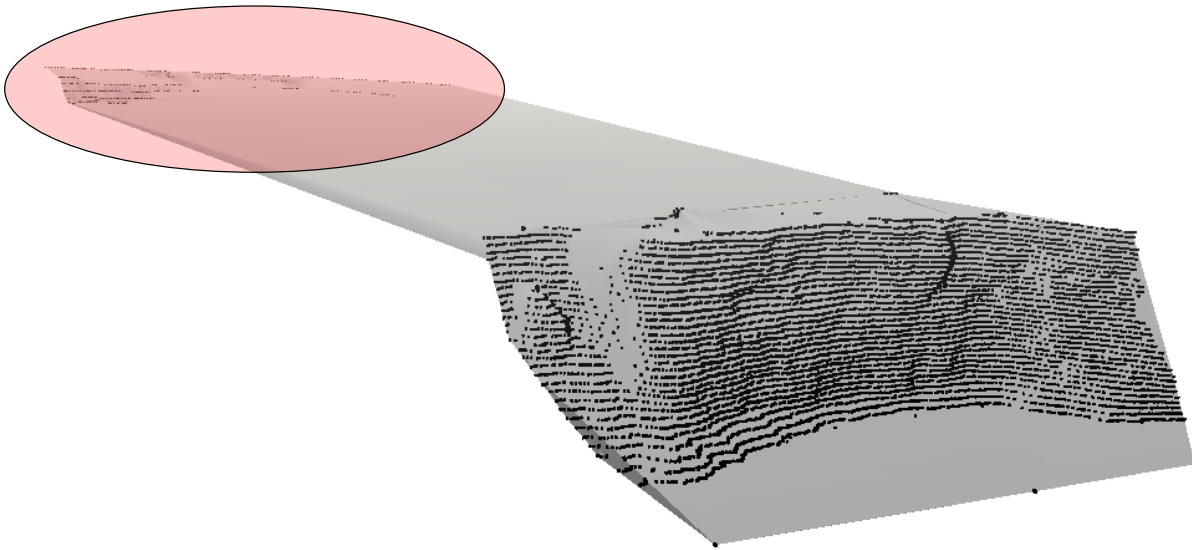


Figura 26 - Pontos atípicos resultados de erro de medição

Vamos aplicar o algoritmo de detecção de *outliers* para resolver esse problema. O primeiro passo é achar o plano que melhor se ajusta aos pontos. Comparando os dois métodos (mínimos quadrados e mínimos resíduos absolutos) podemos comprovar (ou pelos menos intuir) que o método dos mínimos quadrados é muito mais influenciado pelos pontos atípicos.

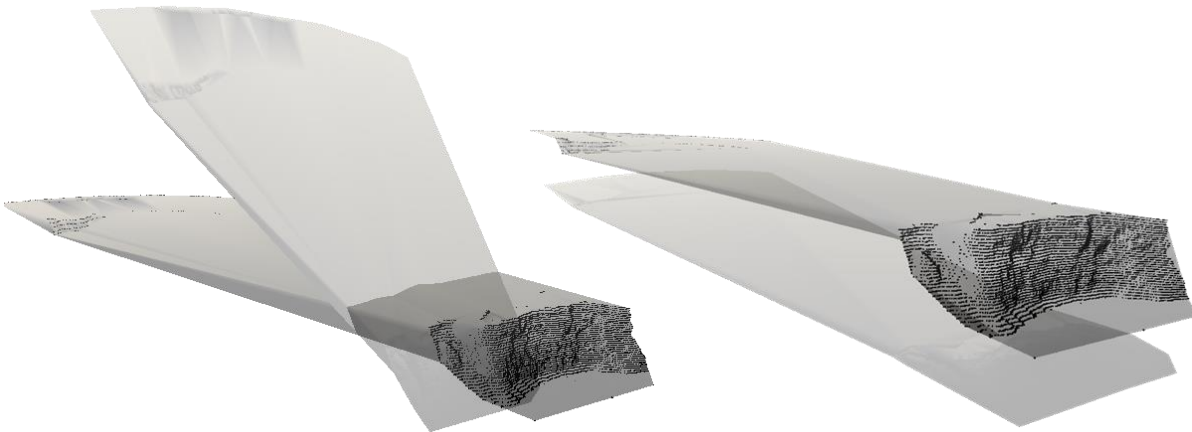


Figura 27 - Hiperplanos de ajuste: mínimos resíduos absolutos (esq.) e mínimos quadrados (dir.).

Usando o método dos mínimos resíduos passamos agora a projetar todos os pontos sobre o hiperplano encontrado.



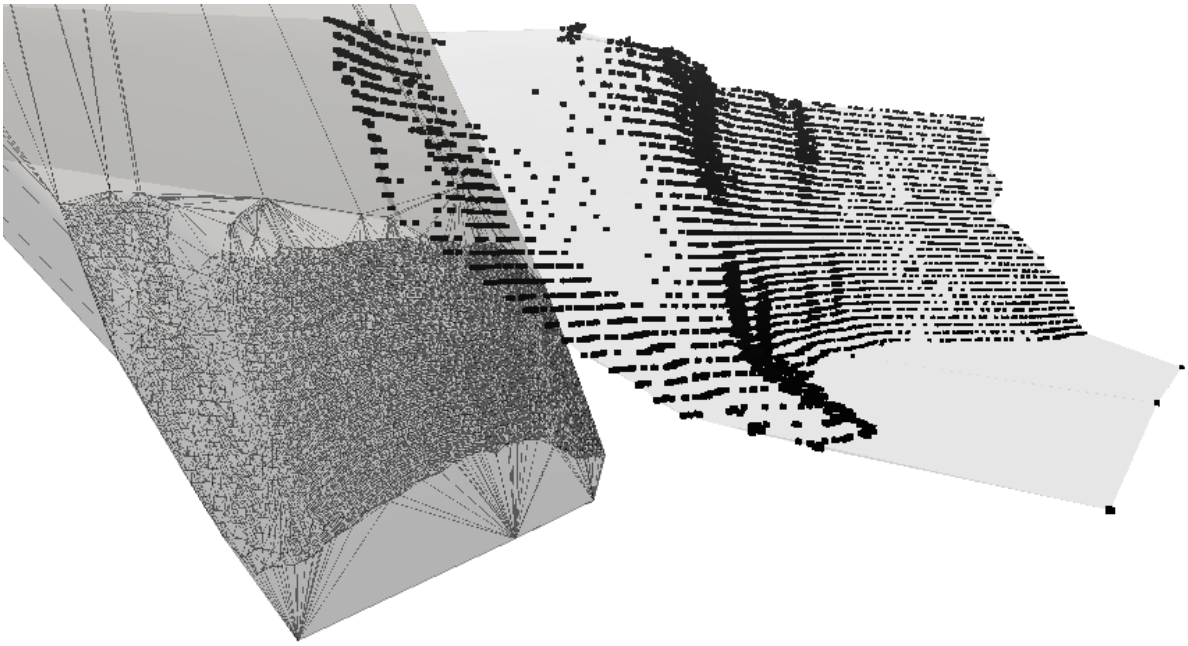


Figura 28 - Projeção dos pontos no hiperplano encontrado.

No exemplo, a média do tamanho dos vetores é 15.01 m, e o desvio padrão é 21.29 m, fixamos um intervalo que compreenda 98% dos dados, assim temos como intervalo:

$$[15.01 - 2.32 \times 21.29 ; 15.01 + 2.32 \times 21.29] = [-34.38 ; 64.40] \quad 37$$

Como as longitudes dos vetores são não negativas, o intervalo se resume a:

$$[0 ; 64.40] \quad 38$$

Assim, os valores que sejam superiores a 64.40m serão eliminados, e temos como resultado:

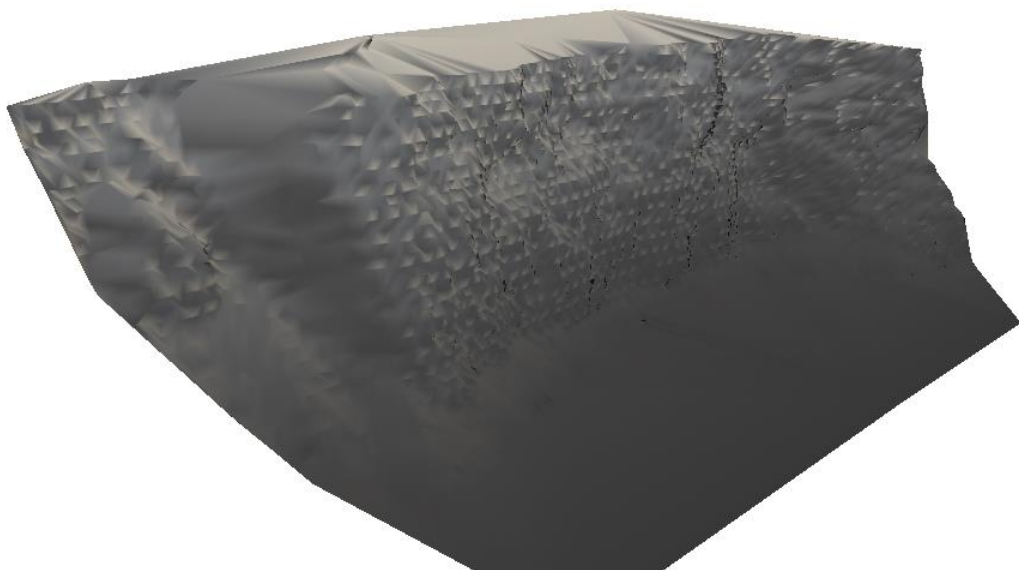


Figura 29 - Modelo topográfico após tratamento para identificar e eliminar *outliers*.

Num próximo exemplo, utilizando fotogrametria, criar um modelo da frente livre:



Figura 30 - Foto de uma frente livre e seu modelo tridimensional gerado através de técnicas de fotogrametria

Observa-se que nesse caso, a presença de plantas no pé da bancada gerou uma nuvem de pontos que, em realidade, não pertencem ao terreno.

Após aplicar a metodologia os pontos atípicos praticamente desaparecem:

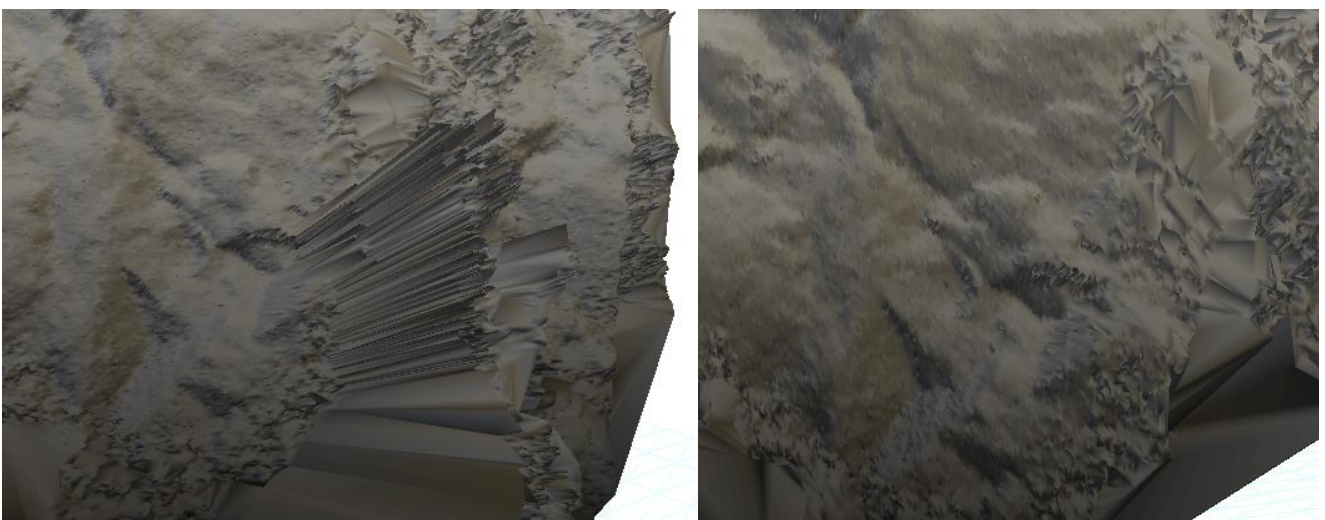


Figura 31 - Terreno com *outliers* gerado pela presença de plantas (esq.) e após a eliminação dos *outliers* (dir.)

## 5

## VALIDAÇÃO DE MODELOS

Quando trabalhamos com regressões lineares, assumimos algumas premissas. Uma delas é assumir a normalidade dos resíduos (Gujarati & Porter, 2010, pág. 99). Um modelo clássico de regressão linear deverá seguir:

$$U \sim N(0, \sigma^2)$$

39

Podemos usar o **teorema do limite central**<sup>9</sup> para justificar teoricamente a normalidade dos resíduos  $u_i$ .

Aqui teremos que comprovar algumas das condições discutidas no apartado 3.3 (“*Suposições assumidas para um modelo de Regressão Linear.*”) ou pelo menos encontrar uma tendência que aponte para essas suposições.

Entenderemos que o modelo está validado se os pontos (ou valores) que desejamos modelar, após um processo de eliminação de *outliers*, gerem projeções no seu hiperplano encontrado mediante técnica de regressão por mínimos resíduos e, após isso, formem vetores entre as projeções e os próprios pontos e consideraremos as longitudes desses vetores (a partir de agora chamados de vetores residuais) como uma variável aleatória que deverá cumprir (ou ter uma tendência clara que aponta ao cumprimento) a condição de ser normal, centrada em zero e com variância constante.

Aplicaremos o exemplo, já utilizado, da triangulação do terreno.

Vamos analisar os vetores residuais **sem** a eliminação dos *outliers* e *a posteriori* aplicando a eliminação.

Para a análise dos dados vamos contar com a ajuda do software SPSS (*Statistical Package for the Social Sciences*), um produto com mais de 30 anos de existência e provável pacote estatístico mais difundido a nível mundial (Pérez, 2001).

Para a tomada de dados (que *a priori* estão serializados em formato binário) é necessário utilizar um software que transforme esses dados para a base decimal e que possamos realizar as transformações necessárias. Como se pode observar na Figura 32, os dados brutos são ilegíveis.

O software escolhido para isso foi o O-Pitblast®, pois conta com todas as ferramentas necessárias para realizar a extração desses dados (Figura 33).

---

<sup>9</sup> A primeira referência aponta para o matemático francês *Abraham de Moivre*, num artigo publicado em 1733. O teorema demonstra que independente da natureza estatística de uma população, se escolhermos amostras de tamanho  $n$ , quanto maior for  $n$  mais a média das amostras se aproximará a uma normal  $\sim (\mu, \sigma^2/n)$  (Samaniego, 2014, pág. 239)

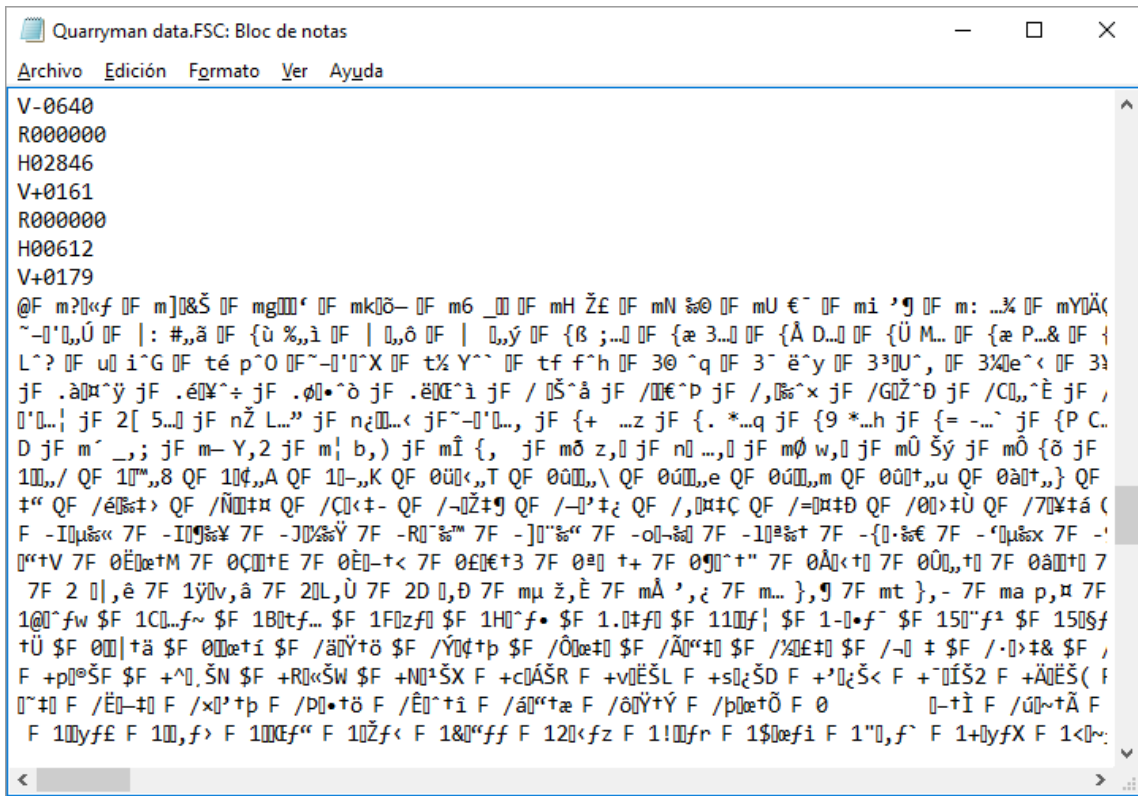


Figura 32 - Exemplo de arquivo gerado por perfilometro 3D antes de qualquer tratamento.

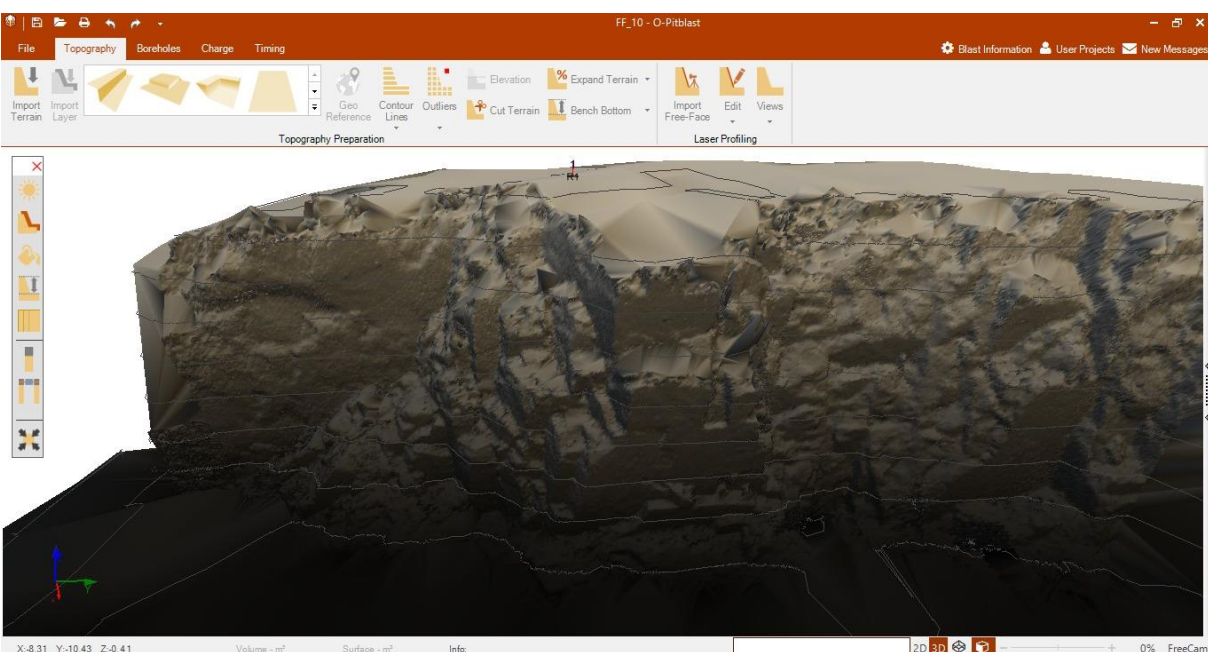


Figura 33 - Tela do Software O-Pitblast®, após interpretar os dados do perfilometro 3D

Ao analisar os dados no SPSS, temos alguns estatísticos:

**Descriptivos**

		Estadístico	Error estándar	
VetoresResiduais	Media	13,2372	,17261	
	95% de intervalo de confianza para la media	Limite inferior	12,8989	
		Limite superior	13,5756	
	Media recortada al 5%	10,0577		
	Mediana	10,1023		
	Varianza	277,469		
	Desviación estándar	16,65739		
	Mínimo	-1,49		
	Máximo	98,10		
	Rango	99,59		
	Rango intercuartil	6,07		
	Asimetría	4,100	,025	
	Curtosis	16,062	,051	

Figura 34 - Estatísticos descritivos

**Pruebas de normalidad**

	Kolmogorov-Smirnov <sup>a</sup>		
	Estadístico	gl	Sig.
VetoresResiduais	,395	9313	,000

a. Corrección de significación de Lilliefors

Figura 35 - Prova de normalidade para os dados com atípicos

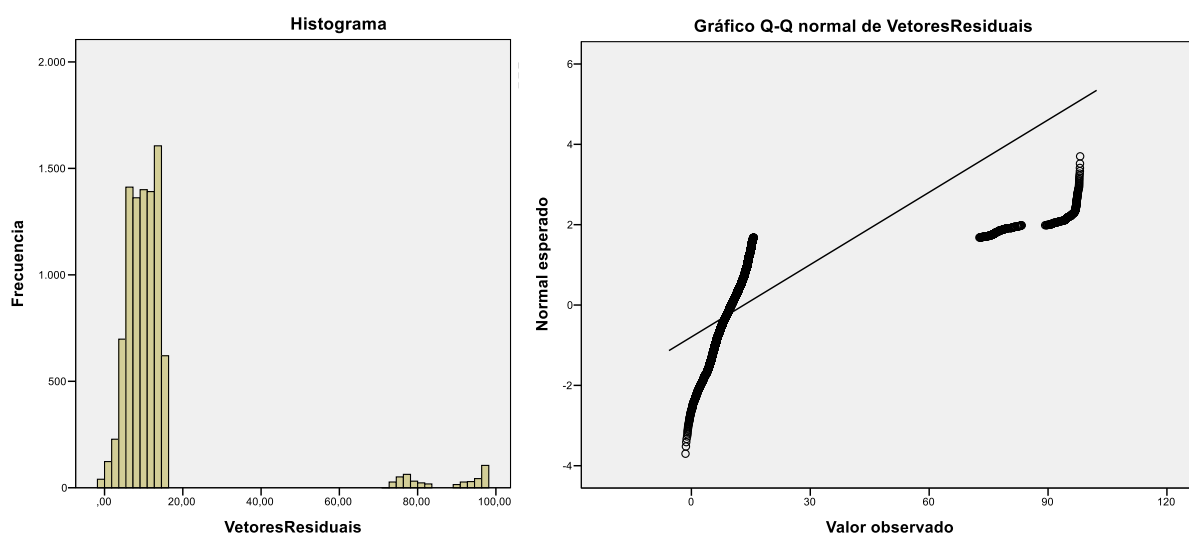


Figura 36 - Histograma e gráfico Q-Q para dados com atípicos

Aqui não fica a menor dúvida da falta de normalidade (estatística) dos dados, pois além da prova de normalidade *Kolmogorov-Smirnov* não aceitar a hipótese nula ( $H_0$ : os dados são normais) em favor da alternativa ( $H_1$ : os dados não são normais), o que não seria conclusivo dado o fato do contraste de hipótese de *Kolmogorov-Smirnov* ser muito rigoroso, mas ao observar o gráfico Q-Q notasse que os pontos não se aproximam da reta, o que descarta qualquer tendência a normalidade. O importante aqui é observar que o histograma nos dá um forte indicativo da presença de *outliers* (Czaplicki, 2014, pág. 59).

Eliminando-se os *outliers* refazemos os testes, e encontramos os resultados:

**Pruebas de normalidad**

	Kolmogorov-Smirnov <sup>a</sup>		
	Estadístico	gl	Sig.
VetoresResiduais	,055	8880	,000

a. Corrección de significación de Lilliefors

Figura 37 - Provas de normalidade para os dados sem atípicos

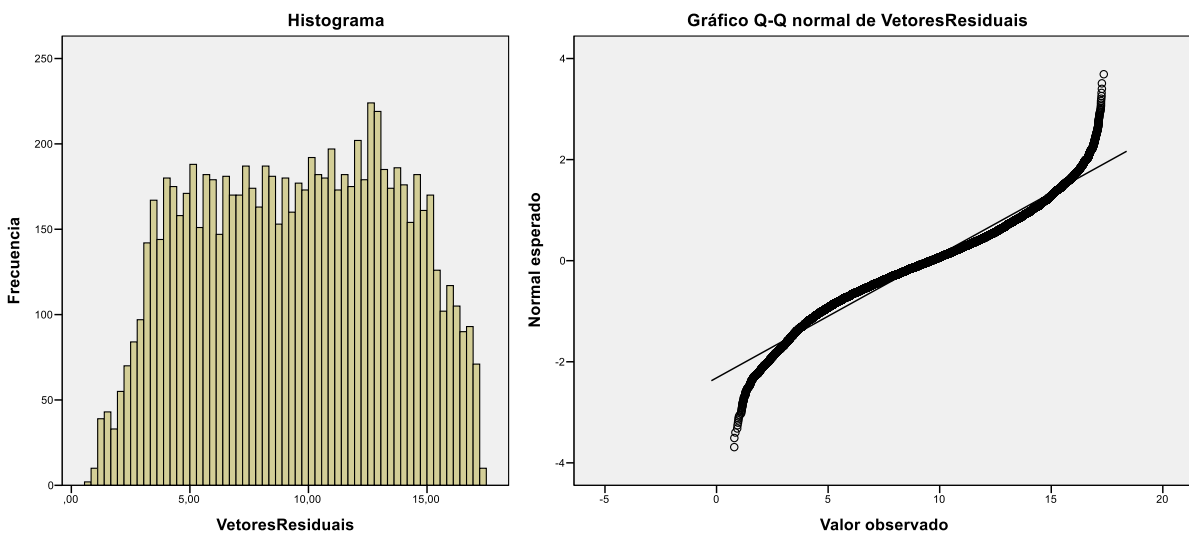


Figura 38 - Histograma e gráfico Q-Q para dados sem atípicos

Aqui podemos observar que os dados também não passam pela prova de normalidade, mas tanto o histograma como o gráfico Q-Q apontam para uma tendência a normalidade.

Para um segundo exemplo vamos utilizar os dados de *Andrade* (2016) e testar a mesma metodologia.

Podemos mais uma vez ver a distribuição dos dados na Figura 39, e aplicamos a metodologia primeiramente detetando os *outliers*.

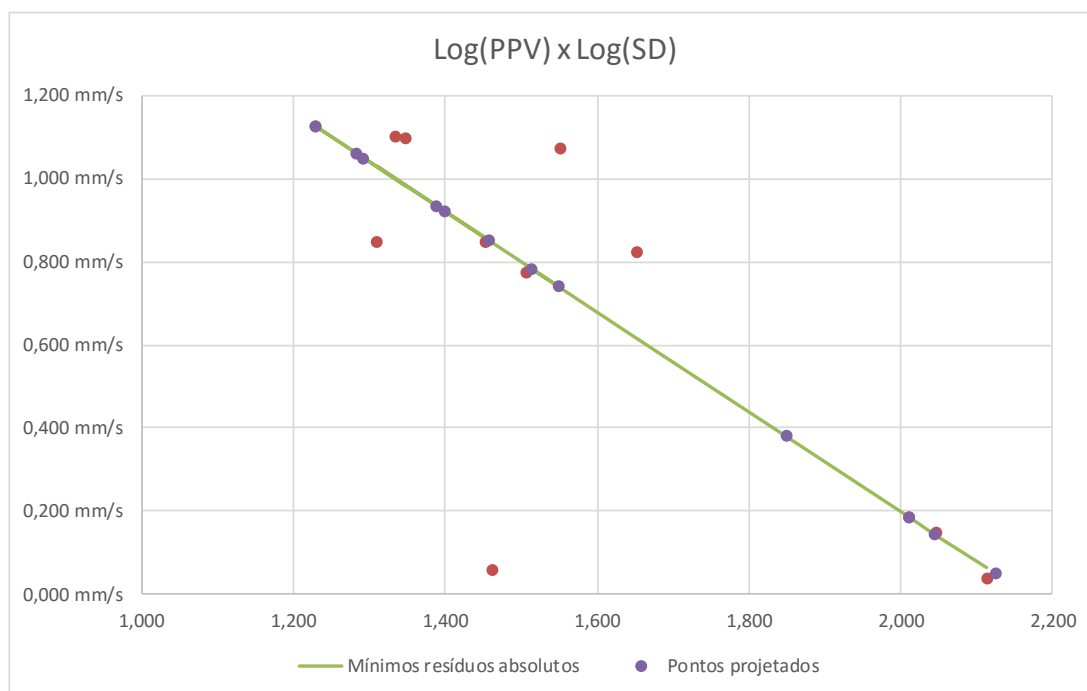


Figura 39 - Log (PPV) x Log(SD), regressão por mínimos valores residuais e vetores projetados

O modelo tem o ponto com menor ordenada eliminado.

Repetimos o teste de normalidade para os vetores residuais.

#### Pruebas de normalidad

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
LogPPV	,268	11	,026	,814	11	,014

a. Corrección de significación de Lilliefors

Figura 40 - Prova de normalidade para os vetores residuais do modelo de previsão do PPV

Aqui também não temos a normalidade confirmada, mas o *p-valor* (“Sig.”) se move e é maior que zero, tanto na prova de *Kolmogorv-Smirnov* quanto no contraste de *Shapiro-Wilk*, o que nos dá indicativo de uma tendência normal (a quantidade pequena de pontos -comprovada pela reprovação no contraste F de Fisher – atrapalha a comprovação da normalidade).

Como se pode ver na Figura 41, o contraste de *Durbin-Watson* se aproxima a dois, o que nos indica a independência dos resíduos (em conformidade com o apartado 3.3.5).

A variância constante é matematicamente difícil de comprovar, sem existir (ao menos baixo os conhecimentos do autor desse trabalho) um contraste de hipótese para essa finalidade. Mas podemos observar na Figura 42 - Diagrama de dispersão dos resíduos que a dispersão dos resíduos não apresenta um padrão ou tendência, o que nos indica fortemente a constância da variância.

Resumen del modelo <sup>b</sup>

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación	Durbin-Watson
1	,944 <sup>a</sup>	,890	,878	,14388	2,129

a. Predictores: (Constante), LogSD

b. Variable dependiente: LogPPV

Figura 41 - Resumo do modelo com contraste de *Durbin-Watson*

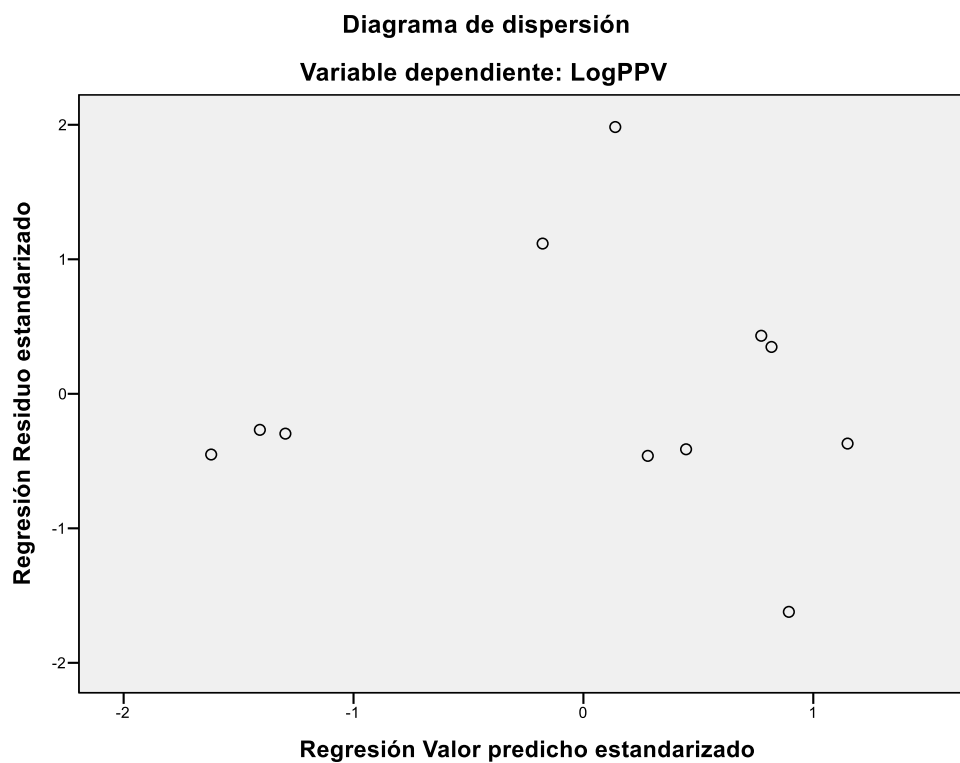


Figura 42 - Diagrama de dispersão dos resíduos



# 6

## **CONCLUSÃO E TRABALHOS FUTUROS**

Podemos concluir que os modelos lineares estão presentes no cotidiano do profissional da Engenharia de Minas e que suas aplicações são importantes para o desenrolar bem-sucedido desse profissional. Fica claro também com esse trabalho que a toma de dados atípicos pode mudar completamente a análise dos resultados inferidos pelo engenheiro e com isso as possibilidades de decisões errôneas são existentes. No entanto, a metodologia proposta indica uma direção para a comprovação dos modelos utilizados pelo engenheiro e um tratamento de detecção e eliminação de *outliers* que permite ao profissional em questão saber, ou pelo menos intuir, quando seu modelo poderá ou não ser confiável.

Para futuros trabalhos seria uma mais valia poder validar a metodologia sugerida com mais casos de campo e outras aplicações lineares que estão presentes no dia-a-dia do profissional da Engenharia de Minas, e aqui citamos:

- Aplicações lineares em teoria de grafos para cálculo de afastamentos críticos;
- Detecção de atípicos em aplicações ARIMA para modelos de previsão;
- Otimização de fluxos através de modelos de redes;
- Outras aplicações.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABNT, A. B. (2005, Set). NBR 9653: Guia para avaliação dos efeitos provocados pelo uso de explosivos nas minerações em áreas urbanas. Rio de Janeiro, Brasil.
- Bhandari, S. (1997). *Engineering Rock Blasting Operations*. Rotterdam: Balkema.
- Boulos, P. (2005). *Geometria Analítica: um tratamento vetorial*. São Paulo: MAKRON Books do Brasil.
- Corporation, G. D. (2016, 07 02). *Gams*. Retrieved from <https://www.gams.com/>
- Czaplicki, J. (2014). *Statistics for Mining Engineering*. Leiden: CRC Press/Balkema.
- Dantzig, G. (1963). *linear Programming and Extensions*. Princeton: Princeton University Press.
- Downing, C. (1985). *Blast Vibration Monitoring and Control*. Ottawa: Prentice-Hall.
- Galton, F., & Hamilton Dickson, J. (1886). Family Likeness in Stature. *Royal Society of London*, 42-73.
- Gass, S. I., & Assad, A. (2005). *An Annotated Timeline of Operations Research: An Informal History*. Boston: Kluwer Academic.
- Goldberg, M. C., & L. Luna, H. P. (2000). *Otimização Combinatória e Programação Linear: modelos e algoritmos*. Rio de Janeiro: Campus.
- Goldberger, A. (1991). *A Course in Econometrics*. Cambridge: Harvard University Press.
- Gouveia de Miranda, V. (2012). Límites Óptimos Para la Toma de Decisiones en la Construcción de Nuevas Oficinas: Un Enfoque Logístico. Madrid: Universidad Complutense de Madrid.
- Grubbs, F. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics* 11, 1-21.
- Gujarati, D., & Porter, D. (2010). *Econometria*. México: McGraw-Hill/Interamericana Editores, S.A. de C.V.
- Hamming, R. (1971). *Introduction to Applied Numerical Analysis*. New York: Mcraw-Hill.
- Hamming, R. (1973). *Numerical Methods for Scientist and Engineers*. New York: McGraw-Hill.
- Hillier, F. S., & Lieberman, G. J. (2010). *Introducción a la Investigación de Operaciones*. Mexico: MCGRAW-HILL / INTERAMERICANA DE MEXICO.
- IBM. (2016, 07 02). *CPLEX Optimizer*. Retrieved from <https://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/>

- Institute, S. (2016, 07 02). *SAS: The Power to Know*. Retrieved from [http://www.sas.com/pt\\_pt/home.html](http://www.sas.com/pt_pt/home.html)
- Lipschutz, S. (1994). *Álgebra Linear: teoria e problemas*. São Paulo: Makon Books.
- Mathews, J., & Fink, K. (2000). *Métodos Numéricos con Matlab*. Madrid: Pearson Educación, S.A.
- Microsoft. (2016, 07 02). *XL: O Solver limites de restrições e células ajustáveis*. Retrieved from <https://support.microsoft.com/pt-br/kb/75714>
- Moreira Andrade Gomes, J. P. (2016). *Metodologia para Análise de Vibrações Provocadas por Desmonte com Recurso a Explosivos*. Porto, Portugal.
- Novalés Cinca, A. (1993). *Econometría*. Madrid: McGraw-Hill/Interamericana de España, S.A.U.
- Pearson, K., & Lee, A. (1903). On the Laws of Inheritance in Man. *Biometrika*, 357-462.
- Peña, D. (2010). *Análisis de series temporales*. Madrid: Alianza.
- Pérez, C. (2001). *Técnicas Estadísticas con SPSS*. Madrid: Prentice Hall.
- Pukelsheim, F. (1994). The Three Sigma Rule. *The American Statistician*, 1-10.
- Samaniego, F. (2014). *Stochastic Modeling and Mathematical Statistics*. Davis: CRC Press.
- Sanchidrián, J. A., & Muñoz, E. (2000). *Curso de Tecnología de Explosivos*. Madrid: Servicio de Publicaciones de la Fundación Gómez-Pardo.
- Silveira Ferreira, P. J. (2013). *Princípios de Econometria*. Venda do Pinheiro: Rei dos Livros.
- Simmons, B. (2016, 07 02). *Mathwords*. Retrieved from <http://www.mathwords.com/o/outlier.htm>
- Singh, S. (2001). *O livro dos códigos*. Rio de Janeiro: RECORD.
- Siskind, D. (2005). *Vibration From Blasting*. Cleveland: International Society of Explosives Engineers.
- Taha, H. (2008). *Pesquisa Operacional: Uma Visão Geral*. São Paulo: Pearson Prentice Hall.
- Tormos Juan, P., & Lova Ruiz, A. (2003). *Investigación Operativa para Ingenieros*. Valencia: Universidad Politecnica de Valencia.