

**UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO**

**ENGENHARIA DA COMPUTAÇÃO**

**CARLOS KAMAY DE SOUZA ABREU**

**GABRIELLE FERREIRA DE OLIVEIRA**

**PROFESSOR**

**BERNARDO SOTTO-MAIOR PERALVA**



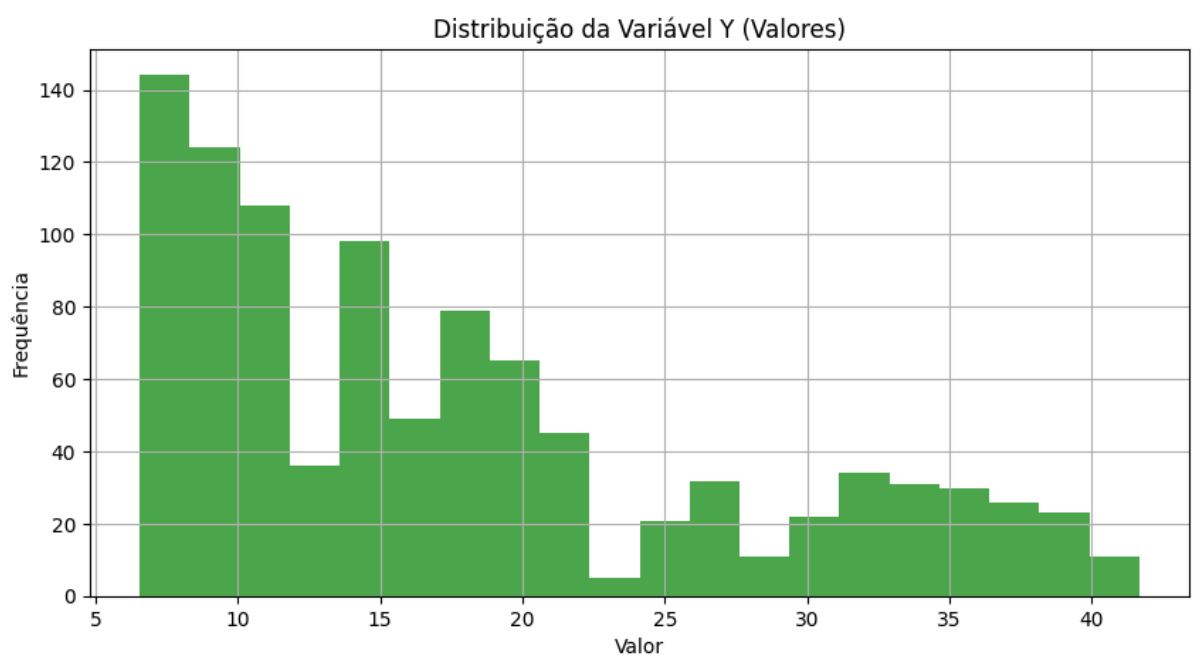
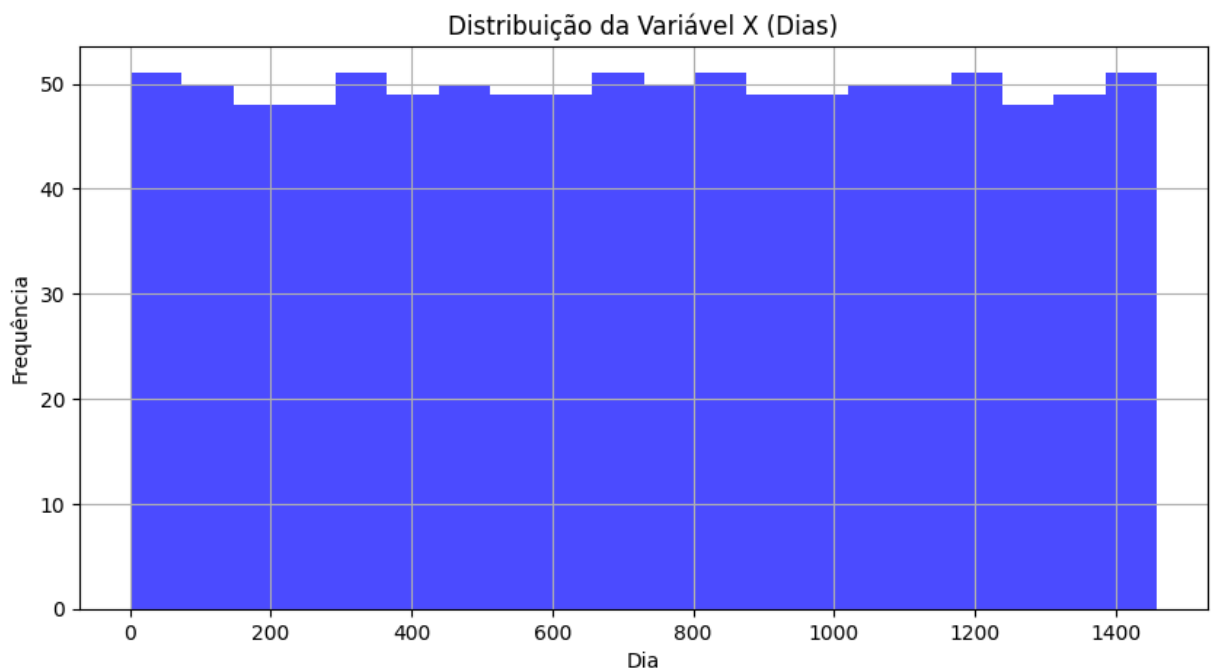
**TRABALHO 1**

**NOVA FRIBURGO**

**2024**

Para realizar esse trabalho utilizamos dados referentes ao valor da ação da Petrobras (PETR4) em um período de mais ou menos 1400 dias. A variável X se refere ao dia e Y ao valor da ação.

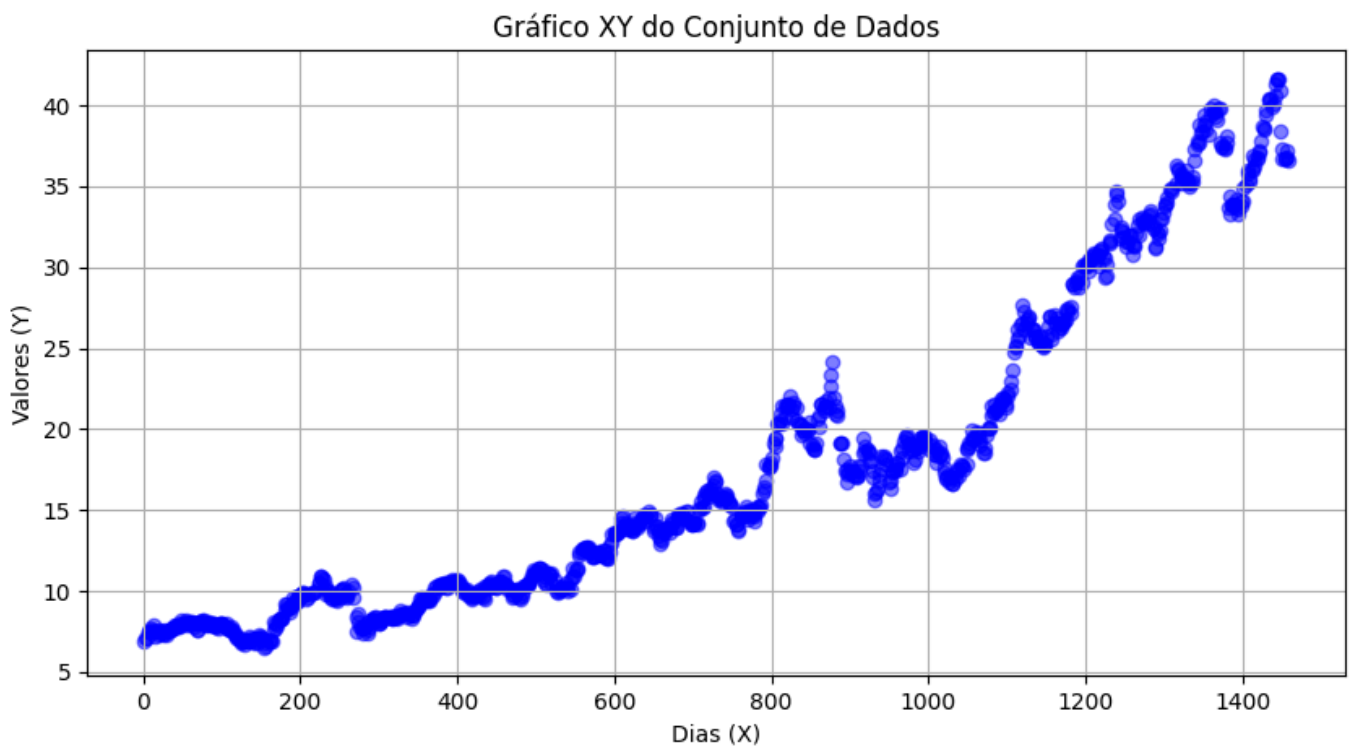
**a) Faça o gráfico das distribuições de cada uma das duas variáveis (Y e X).  
(Dois gráficos separados)**



**b) Através das distribuições é possível identificar pontos influentes em alguma das variáveis? Comente.**

Pela distribuição da variável Y é possível perceber pontos entre 20-25, 25-30 e 40+ que são bem menores do que a média e por isso podem ser considerados pontos influentes.

**c) Faça o gráfico xy do conjunto de dados. Visualmente, é possível identificar alguma correlação entre as variáveis?**



Os pontos no gráfico tendem a subir à medida que se movem para a direita, isso sugere uma correlação positiva.

**d) Calcule o coeficiente de correlação e comente o resultado.**

Média de Y: 17.82

Mediana de Y: 14.95

Desvio Padrão de Y: 9.60

Coeficiente de Correlação: 0.9342

O coeficiente de correlação possui valor próximo de 1, que indica uma forte correlação positiva, sugerindo que, à medida que a variável X(dias) aumenta, a variável Y (valores) também tende a aumentar.

**e) Encontre a reta de quadrados mínimos (estime  $\beta_0$ ,  $\beta_1$  e  $\sigma^2$ )**

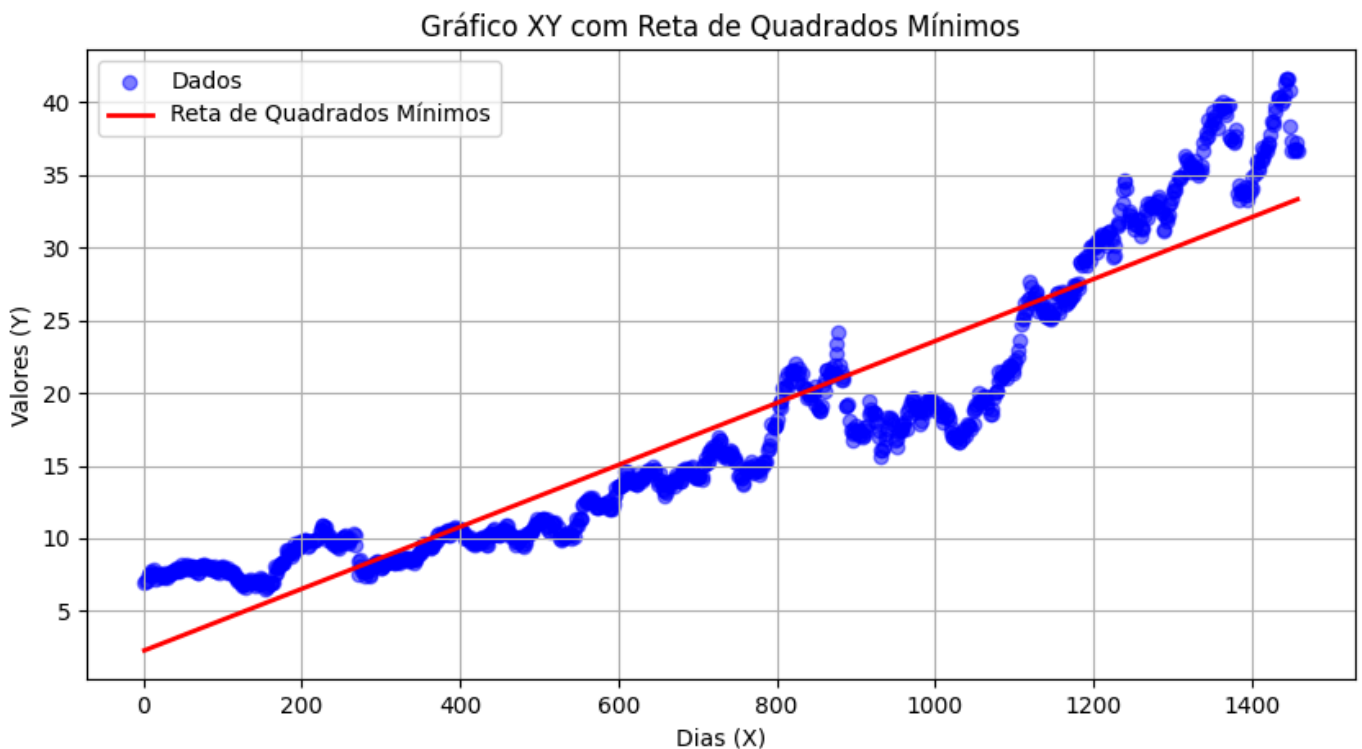
Coeficiente  $\beta_0$ : 2.2720

Coeficiente  $\beta_1$ : 0.0213

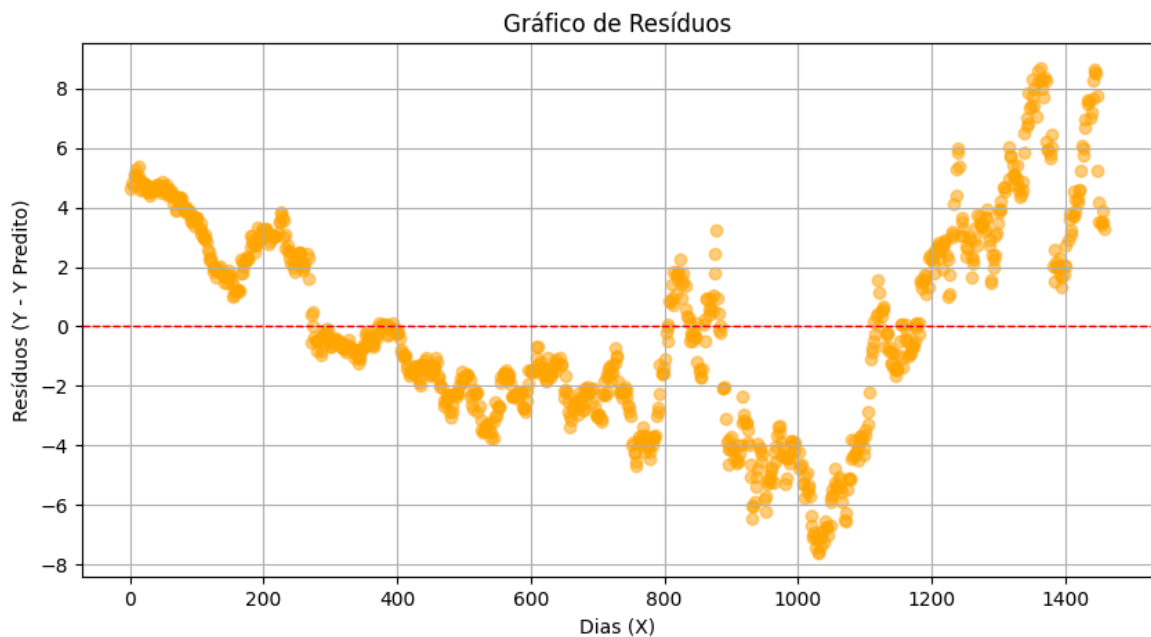
Variância do erro ( $\sigma^2$ ): 11.7357

$Y = 2.2720 + 0.0213.X$

**f) Faça o gráfico os dados no plano xy incluindo a reta de quadrados mínimos encontrada no item anterior.**



**g) Calcule os resíduos.**



Resíduos dos primeiros 10 dados:

Resíduo 1: 4.6367

Resíduo 2: 4.8154

Resíduo 3: 4.7515

Resíduo 4: 5.1002

Resíduo 5: 5.1089

Resíduo 6: 5.0676

Resíduo 7: 5.2863

Resíduo 8: 5.3724

Resíduo 9: 5.0711

Resíduo 10: 4.9298

Visualmente é possível relacionar a grande concentração de resíduos perto da área 1000 e 1400 no gráfico com os pontos de influência anteriormente analisados na questão b). Pois são a área com a maior quantidade de resíduos positivos e negativos.

**i) Monte a tabela ANOVA e, para o intervalo de confiança de 95%, verifique se a hipótese nula (modelo simples) é rejeitada (utilize a tabela da distribuição F do livro).**

Soma dos Quadrados Total (SQTot): 91534.4496

Soma dos Quadrados da Regressão (SQReg): 79892.6798

Soma dos Quadrados dos Resíduos (SQR): 11641.7698

Grau de Liberdade Total (df\_total): 993

Grau de Liberdade da Regressão (df\_regressao): 1

Grau de Liberdade dos Resíduos (df\_residuo): 992

Média dos Quadrados da Regressão (MS\_regressao): 79892.6798

Média dos Quadrados dos Resíduos (MS\_residuo): 11.7357

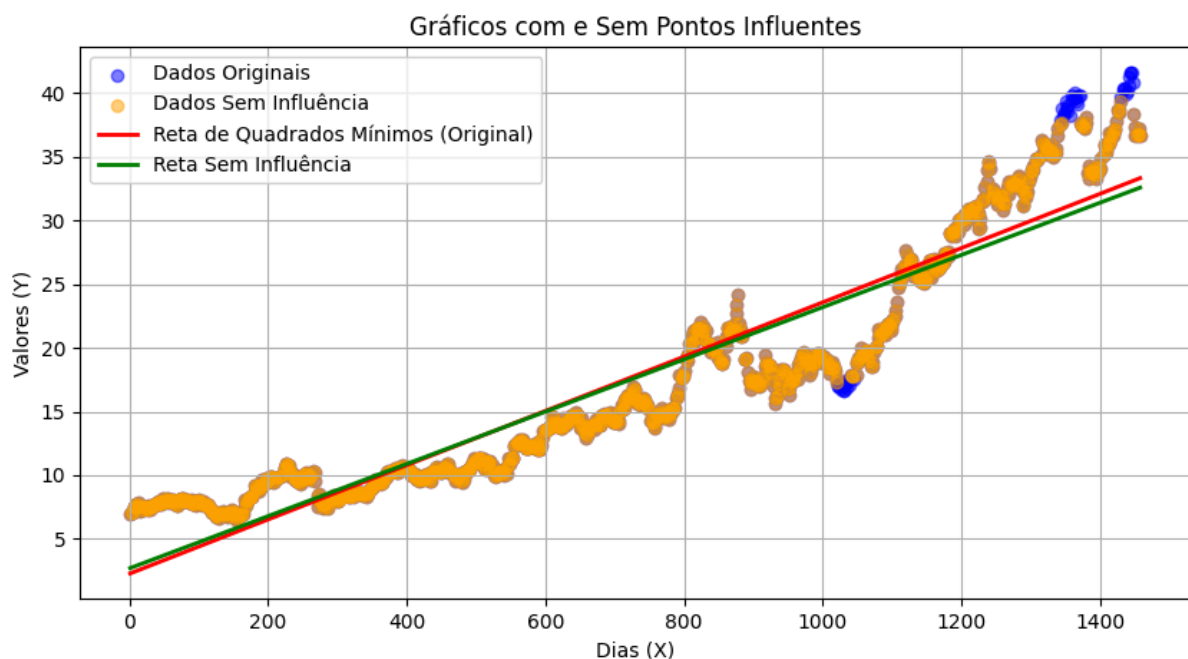
Estatística F: 6807.6881

Valor Crítico F (0.05): 3.8508

A hipótese nula é rejeitada: o modelo linear é significativo.

**j) Retire os pontos classificados como influentes e repita os passos e) e f).**

**Comente sobre os resultados.**



Estimativas com dados sem pontos influentes:

Coeficiente B0 (intercepto) sem influentes: 2.6934

Coeficiente B1 (inclinação) sem influentes: 0.0205

Variância do erro ( $\sigma^2$ ) sem influentes: 9.4227

$$Y = 2.6934 + 0.0205.X$$

A retirada dos pontos influentes interferiu minimamente com a reta de quadrados mínimos, mas ainda assim consegue perceber uma mudança na reta e diminuição da variância do erro.

