WORKSHOP DE REGRESSÕES NÃO LINEARES PARA ENGENHEIROS DE PETRÓLEO (COM PYTHON)

Misael Edgar Zepeda Díaz

mzepeda@petrointelligence.com
(+52 55 1350 7432)

Setembro 2023

PI PETRO Intelligence®

SPE International®

Federal University of Sergipe
SPE Student Chapter

PETROIntelligence®.com

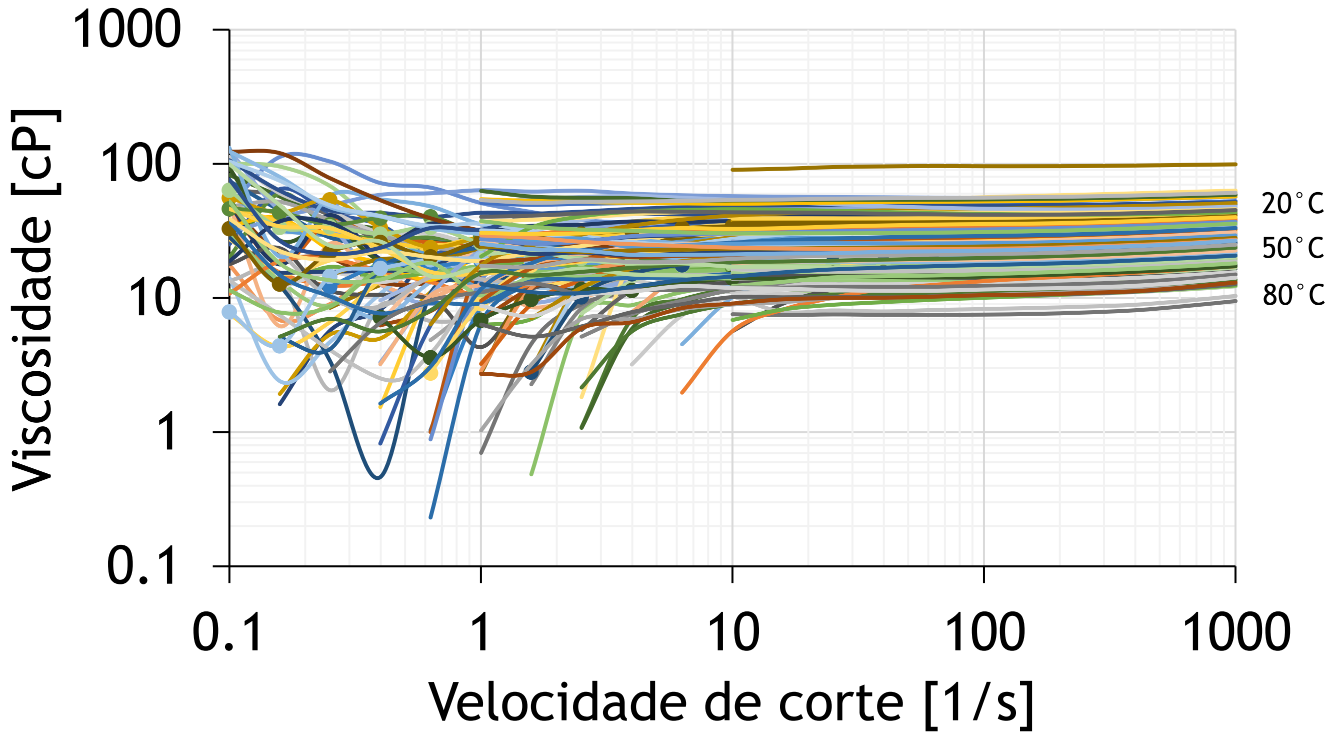
INTRODUÇÃO

Abordagem prática:

1. Coleta, organização e visualização de dados.
2. Análise de dados.
3. Modelado.
4. Regressões lineares e não lineares.
5. Uso de Excel e Python.

Petróleo leve

1. ***Base de dados:***
    a. Viscosidade.
    b. Velocidade de corte.
    c. Temperatura.
    d. Densidade.

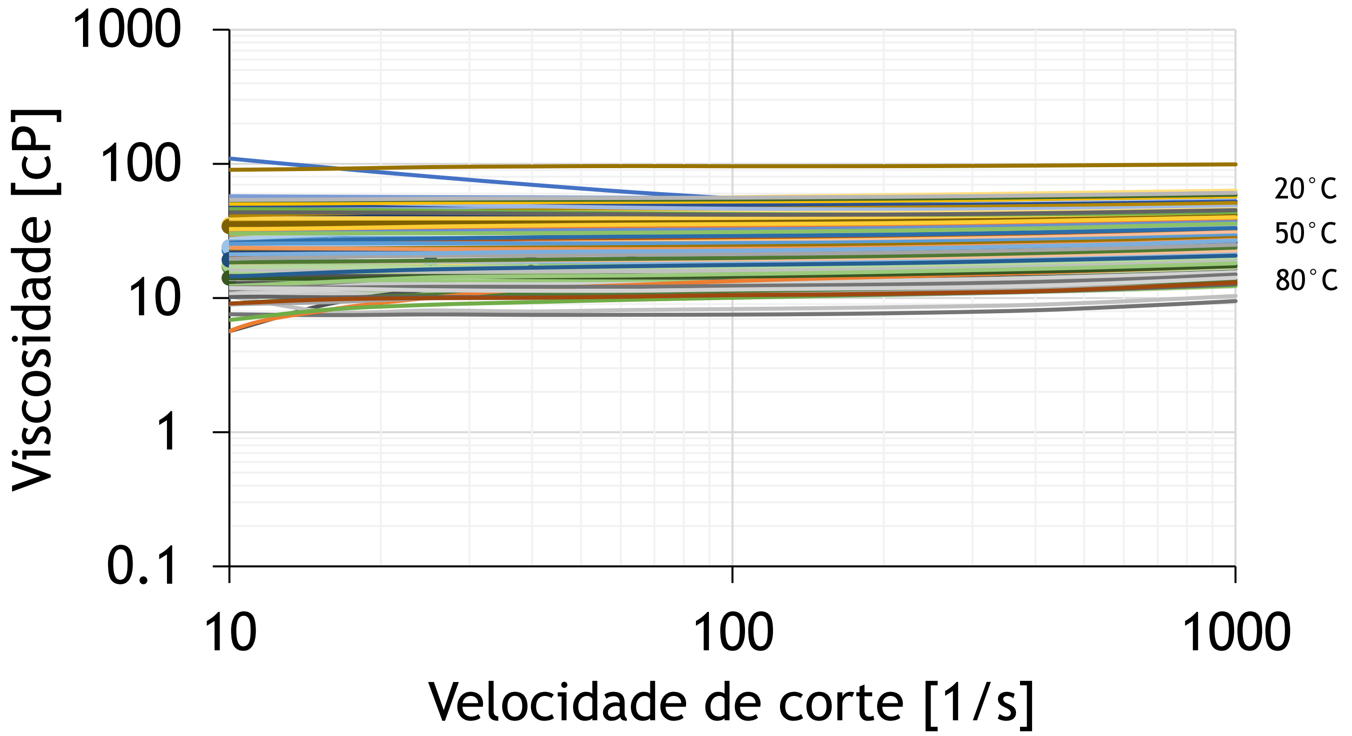


* Zepeda Díaz, M. 2022. Estudo do comportamento reológico de amostras de condensado. Dissertação de Mestrado, UNAM, Programa de Mestrado e Doutorado em Engenharia, Cidade do México.

2. Definição da gama de análise:

a. Faixa de interesse.

b. Tipo de comportamento.

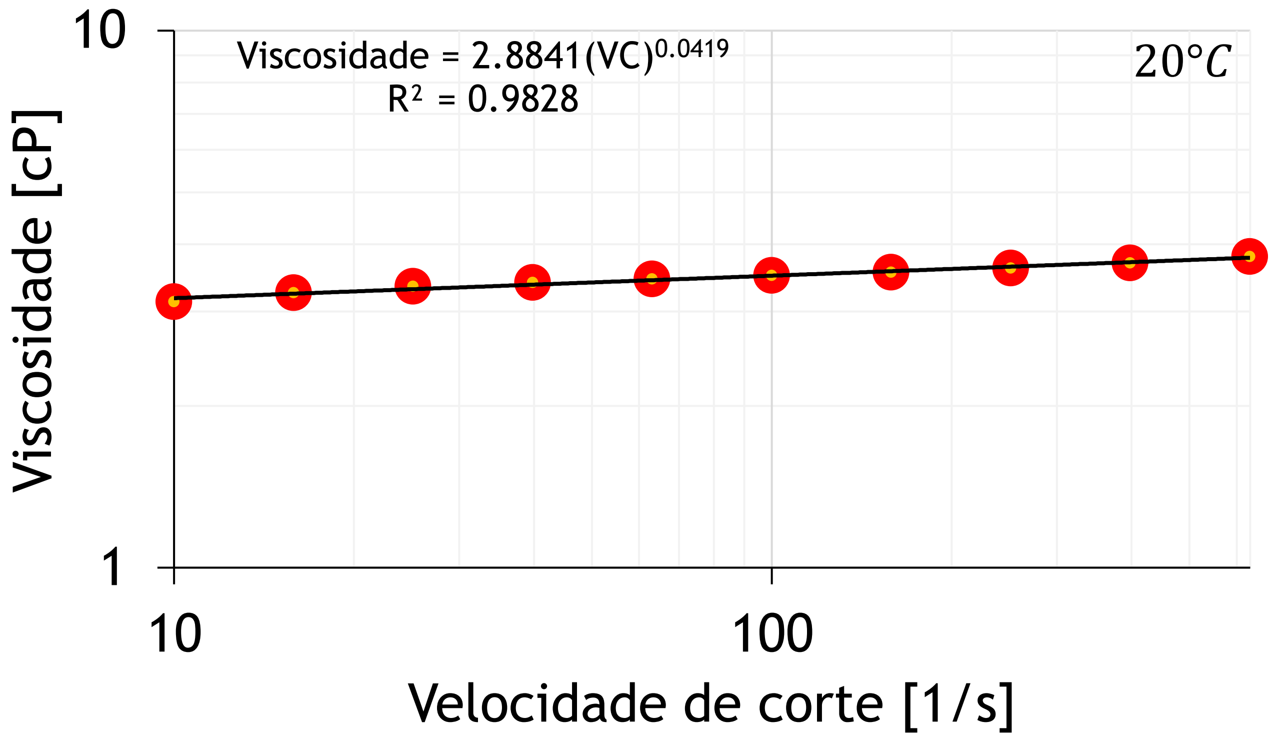


3. Definição do modelo:

$$\eta = A\dot{\gamma}^B$$

$$T, \rho = cte$$

$$\eta = A(\dot{\gamma}^B)(T^C)(\rho^D)$$

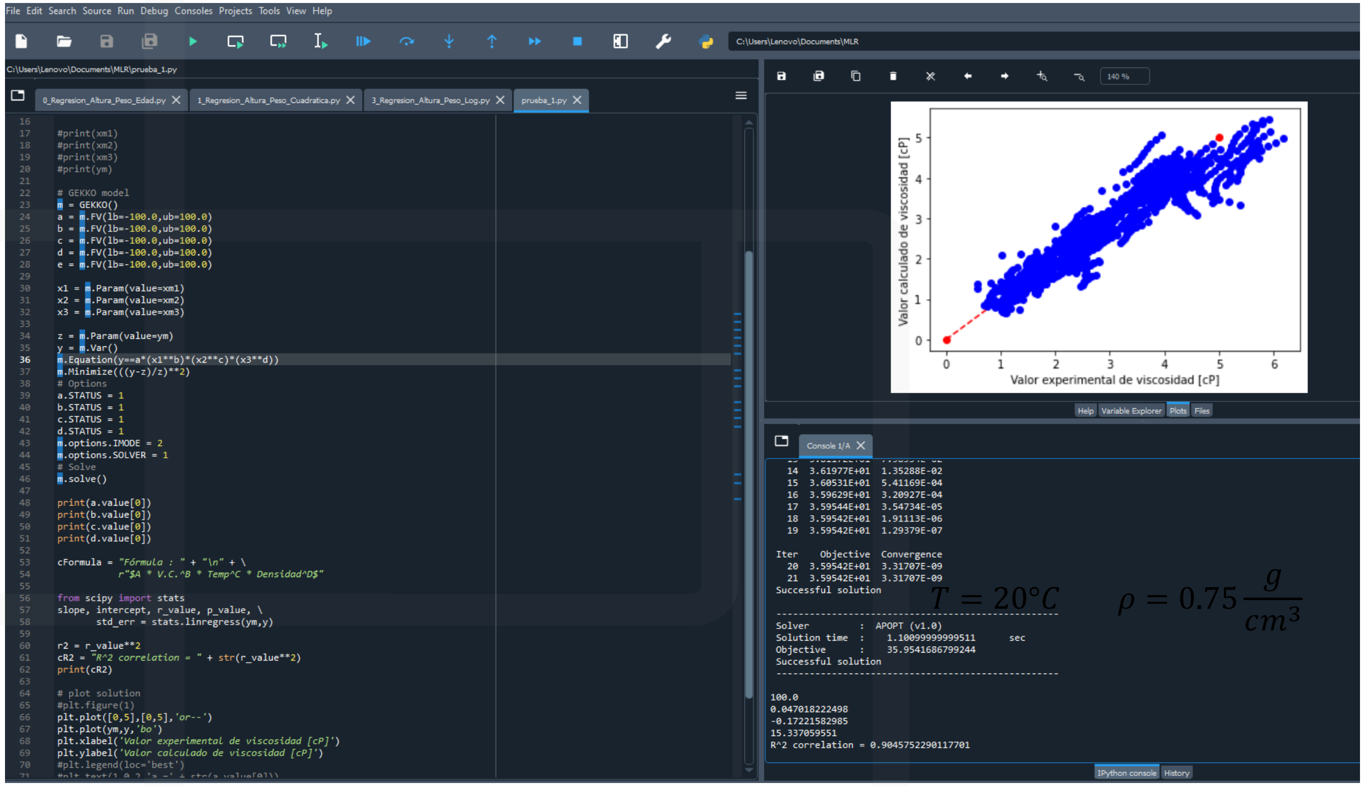


4. Regressão com Python:

a. Coeficientes de regressão.

b. Coeficiente de determinação “R^2”.

c. Avaliação gráfica da correlação.

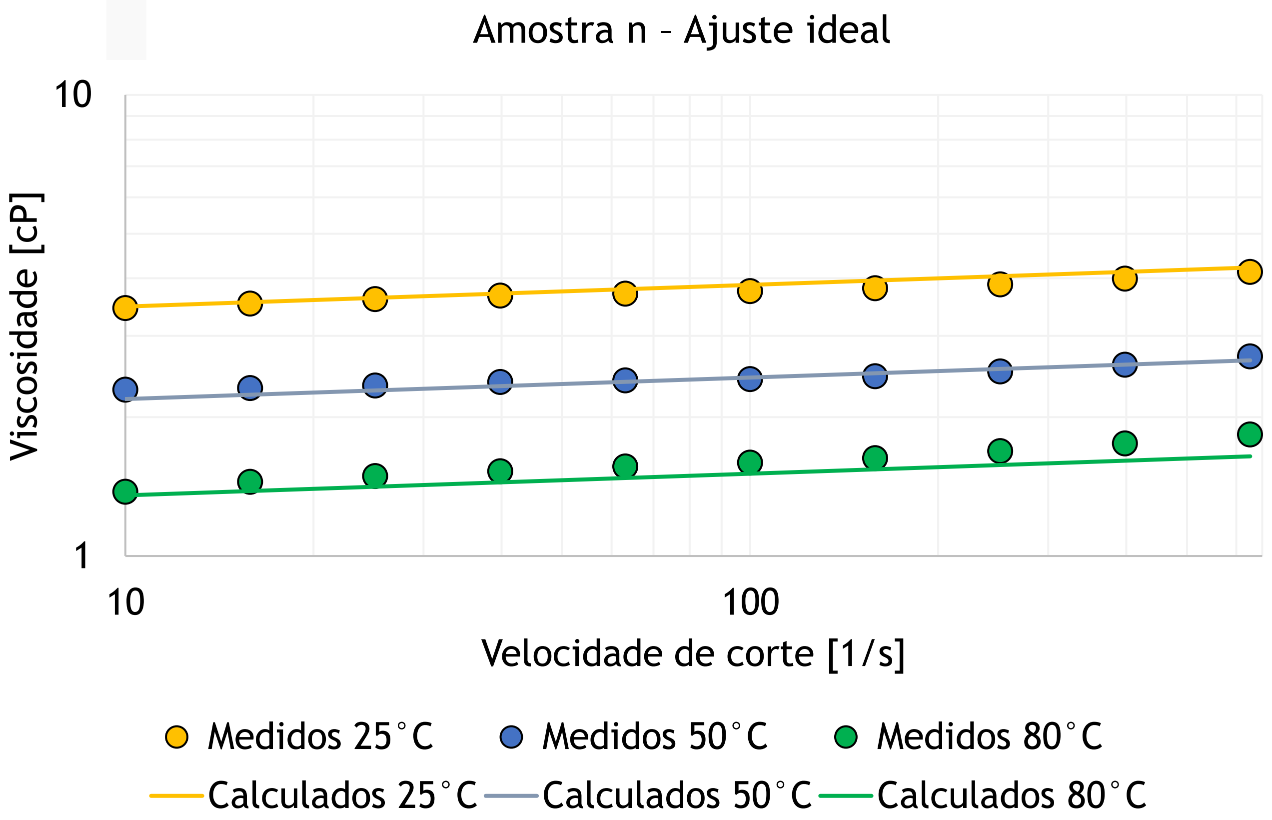


5. Avaliação de regressão:

a. Erro relativo médio.
b. Desvio médio.
c. Gráfico de dados reais vs dados calculados.

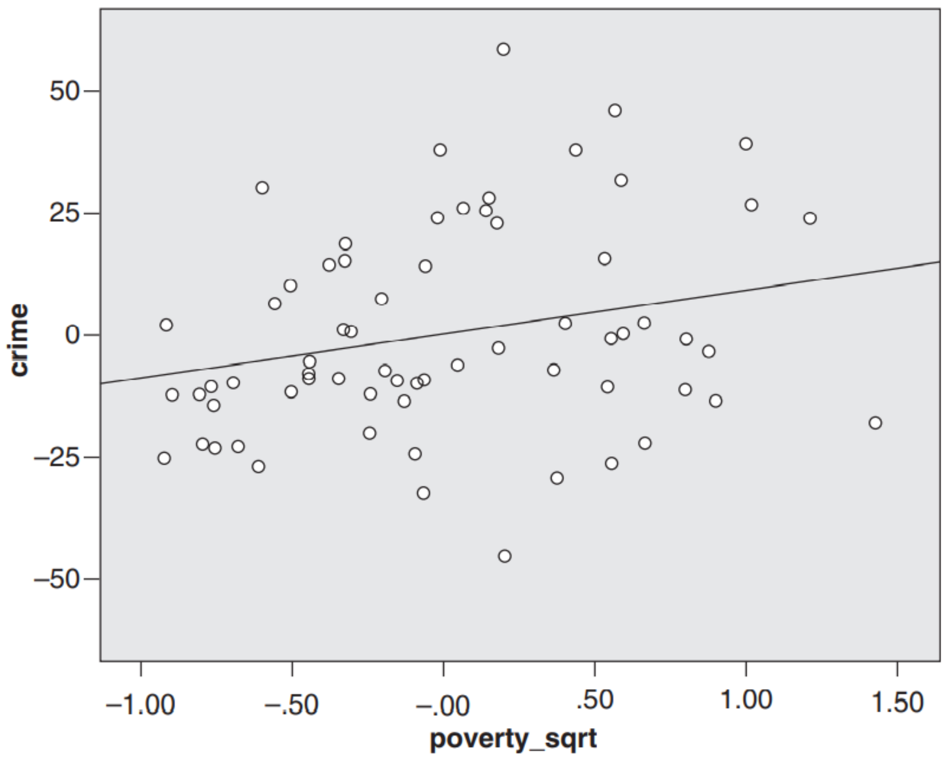
$$ERP_{3c} = 3.31\%$$

$$DP_{3c} = 0.08\ [cP]$$



Regressão:

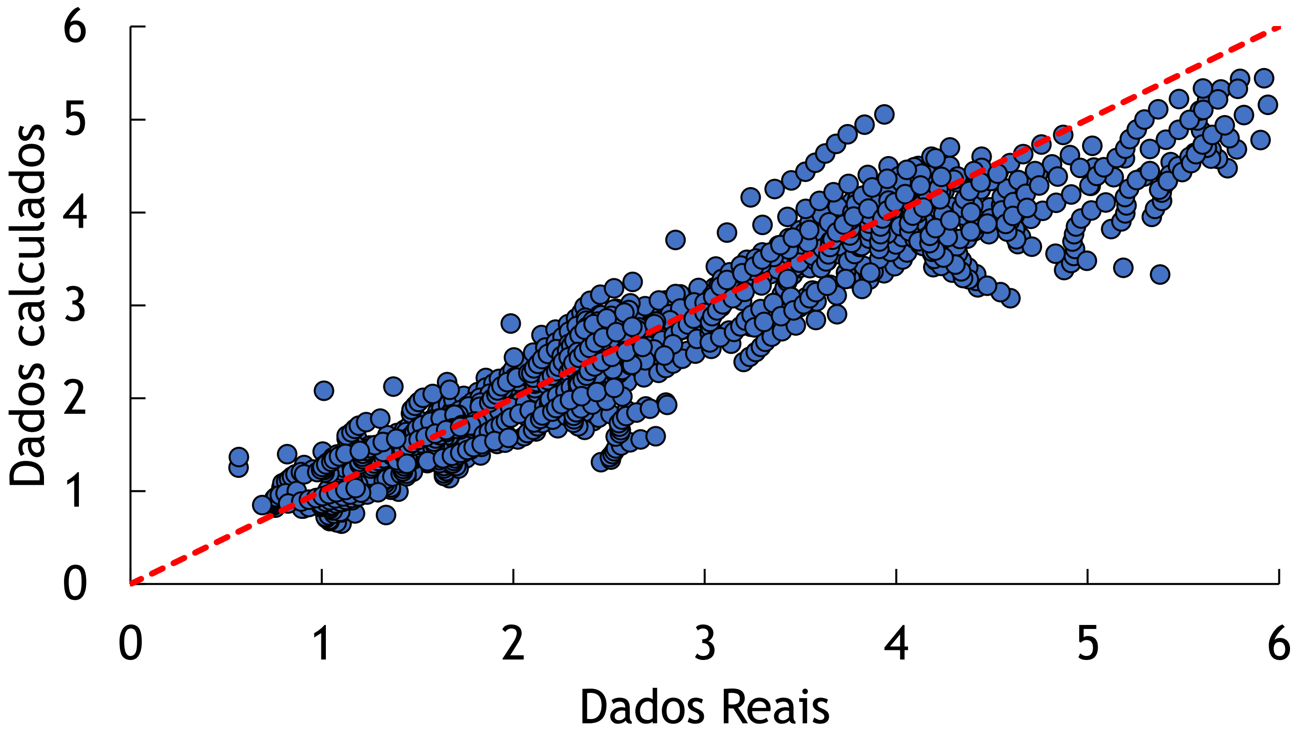
É uma ferramenta estatística para estimar as relações entre uma variável dependente e uma ou mais variáveis independentes.



Yang, Qinghua. (2017). Regression. 10.1007/978-3-319-32001-4_174-1.

Tipos de regressões:

1. Regressão linear.
   a. Simples.
   b. Múltiplo.
2. Regressão não linear.
   a. Simples.
   b. Múltiplo.



1. Regressão linear.

a. Simples.

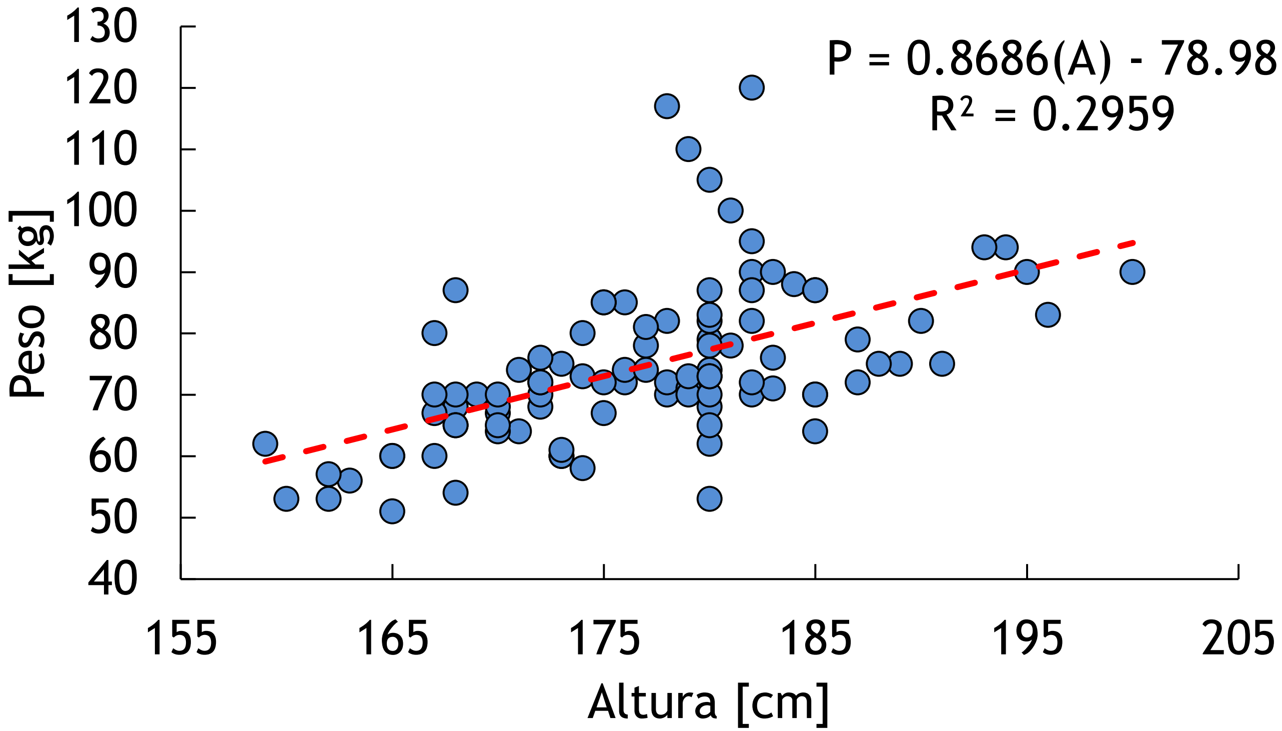
$$y = \alpha + \beta x + \varepsilon$$

y: Variável de destino.

x: Variável independente.

α, β: Coeficientes de regressão.

ε: Erro.



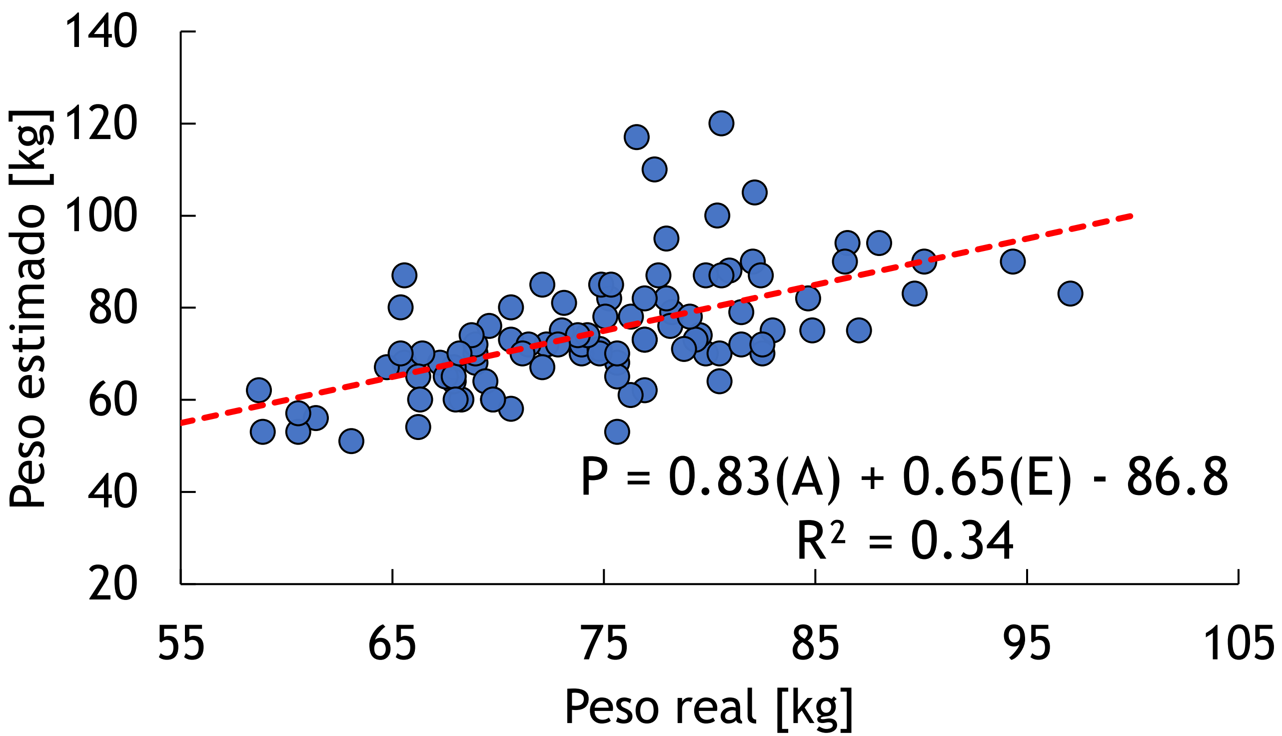
$$\#\boldsymbol{Constantes} = \#\boldsymbol{variables\ independentes} + \mathbf{1}$$

1. Regressão linear.

b. Múltiplo.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

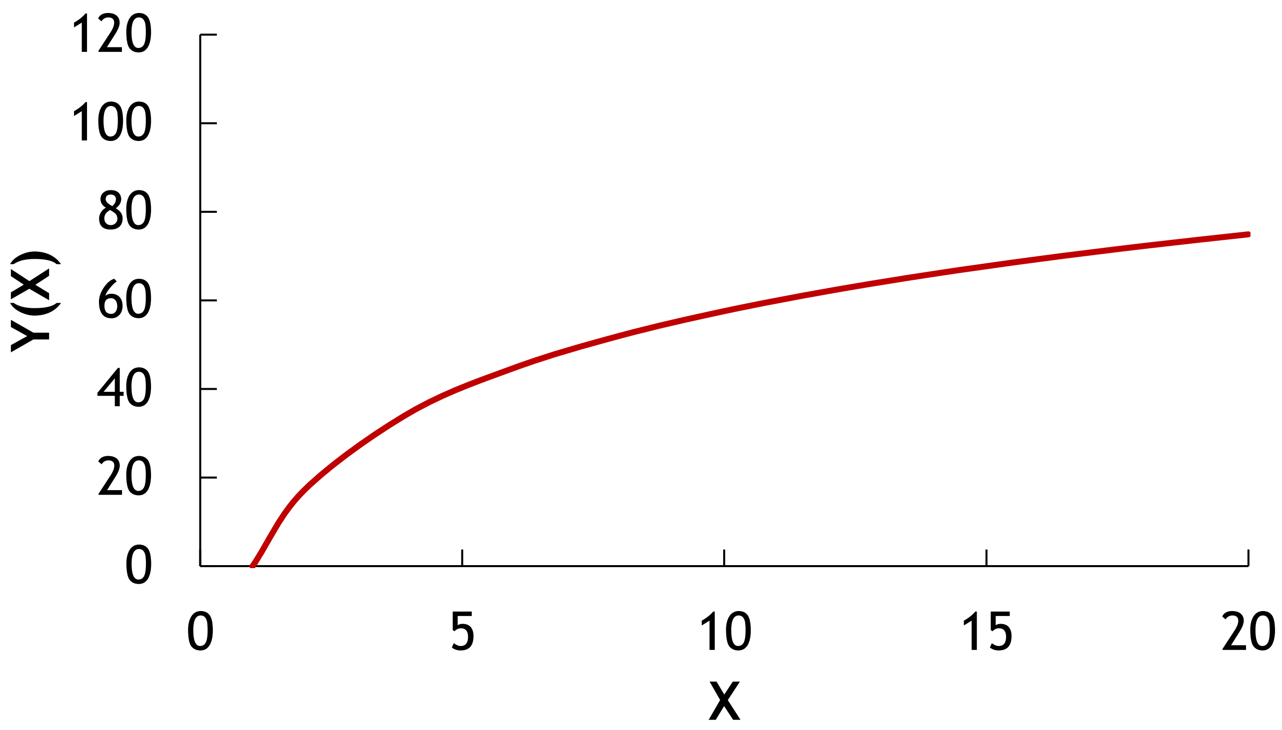
y:	Variável de destino.
$x_1, x_2, .., x_k$:	Variável independente.
$\beta_1, \beta_2, .., \beta_k$:	Coeficientes de regressão.
ε:	Erro.
k:	K-ésimo termo.



2. Regressão não linear.

- Quadrático.
- Potencial.
- Logarítmica.

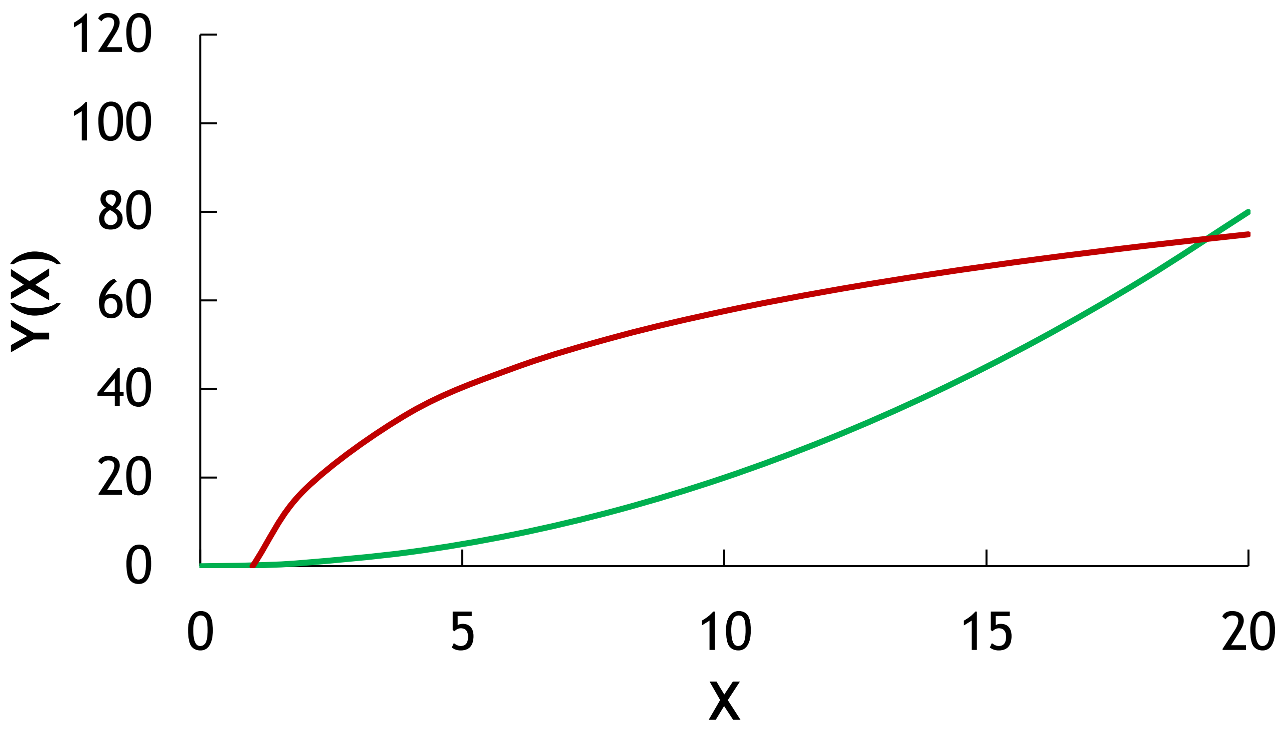
¿?



2. Regressão não linear.

- Quadrático.
- Potencial.
- Logarítmica.

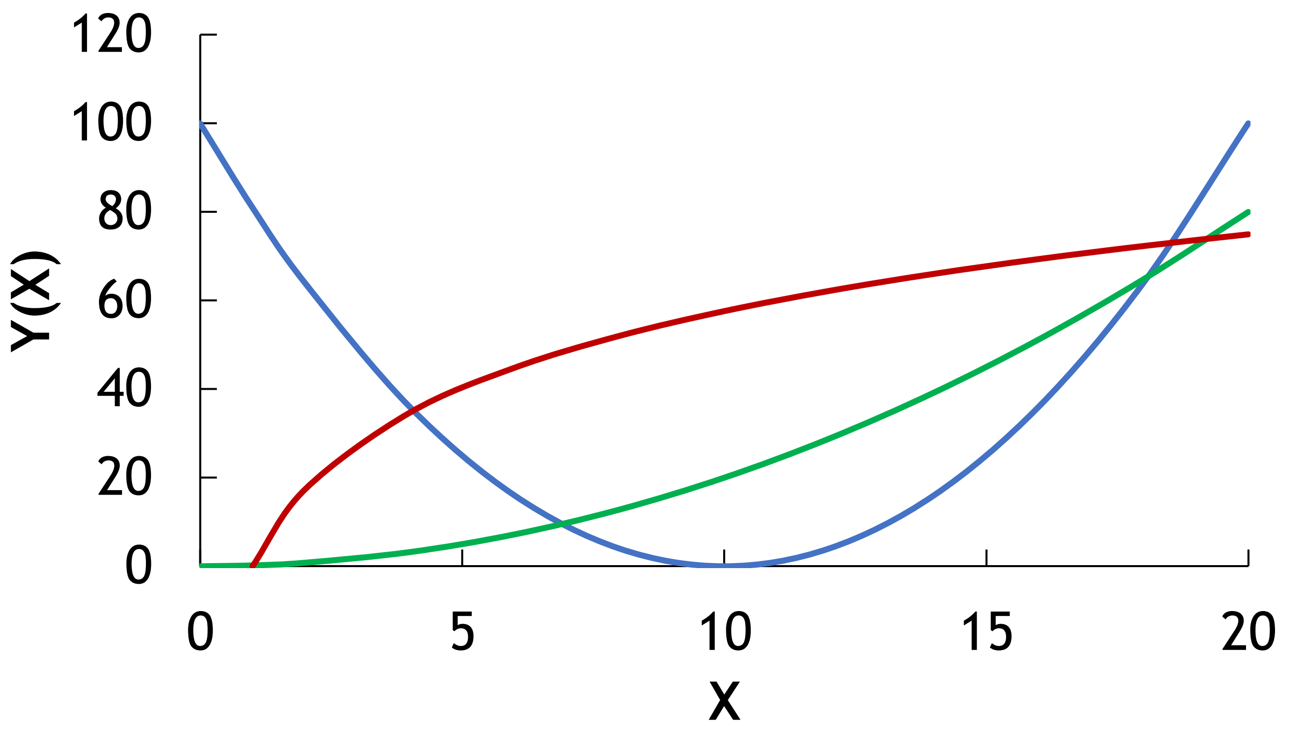
¿?



2. Regressão não linear.

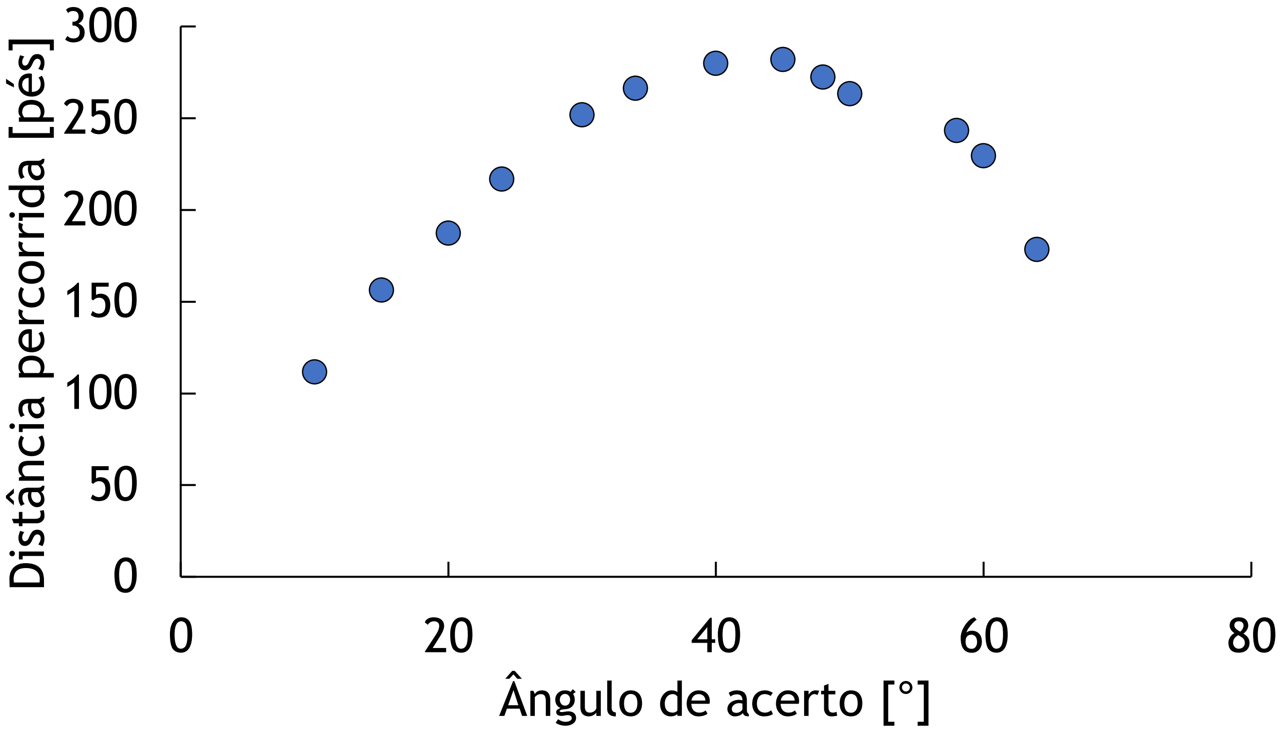
- Quadrático.
- Potencial.
- Logarítmica.

¿?



2. Regressão não linear.

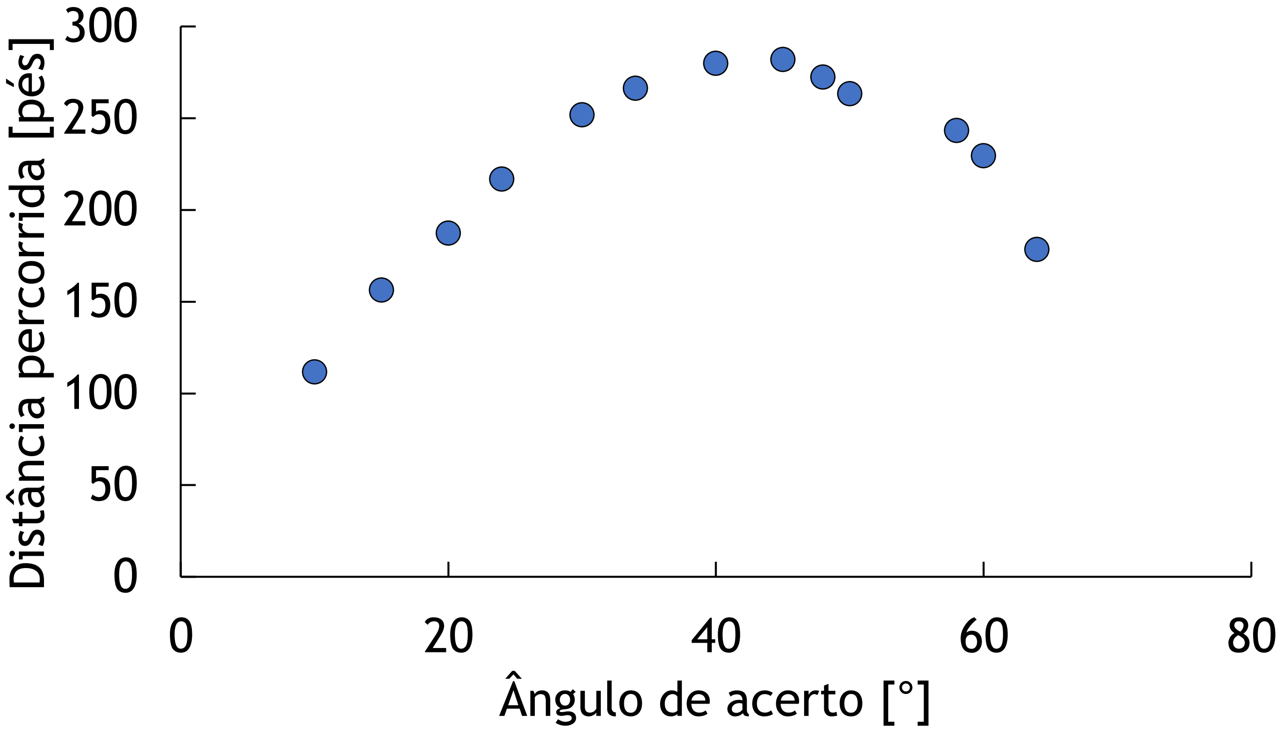
- ¿?



2. Regressão não linear.

- Quadrático.

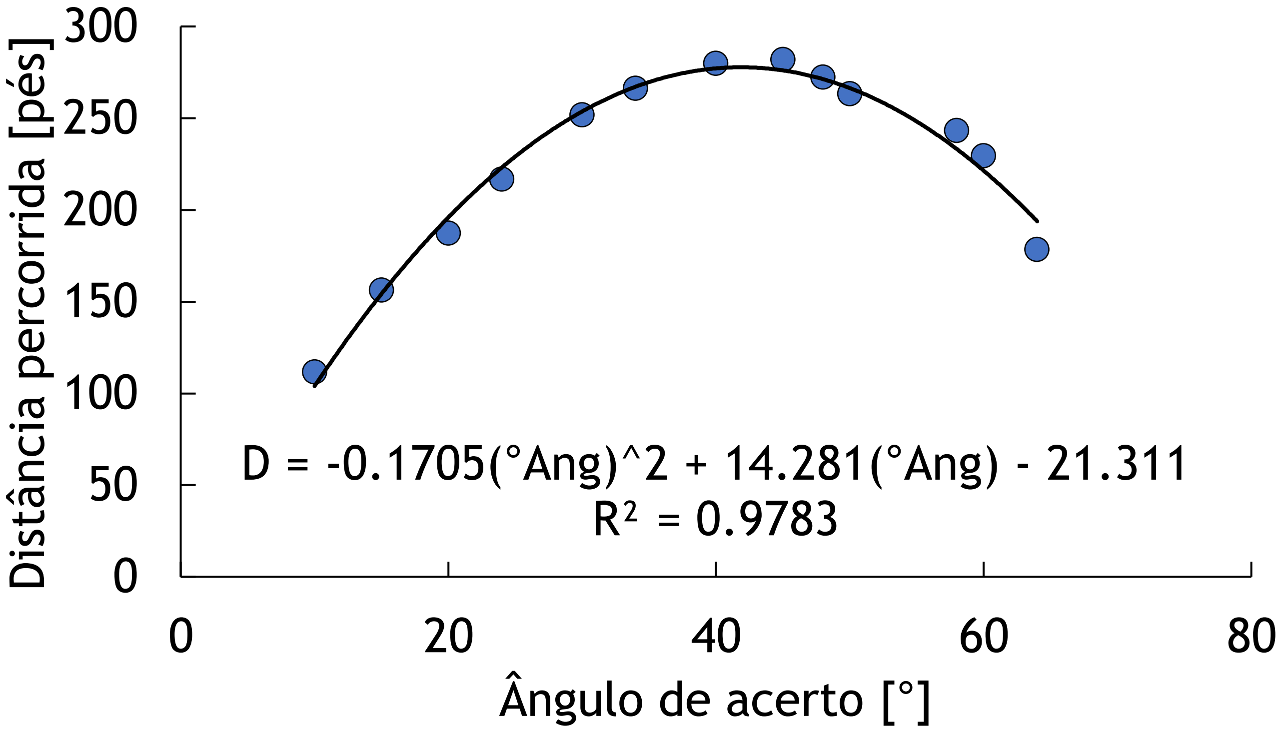
$$y = \alpha + \beta x_1 + \gamma x_1^2 + \varepsilon$$



2. Regressão não linear.

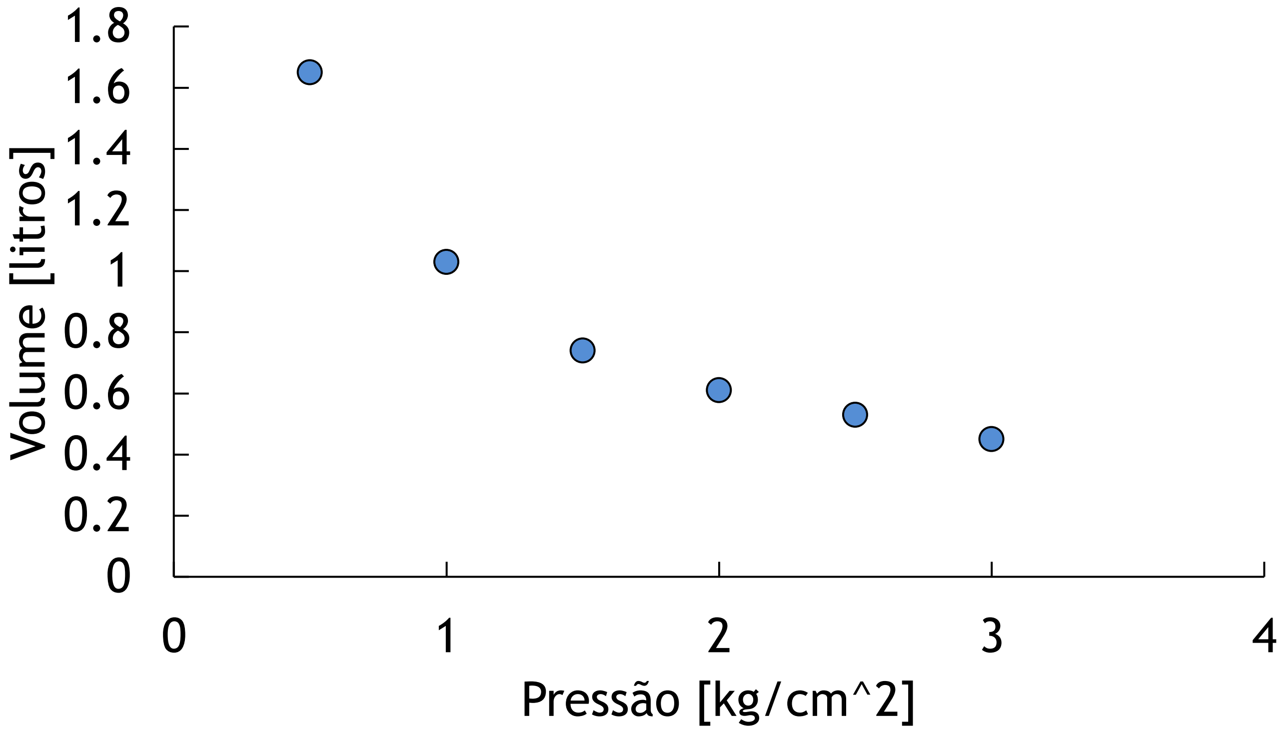
- Quadrático.

$$y = \alpha + \beta x_1 + \gamma x_2^2 + \varepsilon$$



2. Regressão não linear.

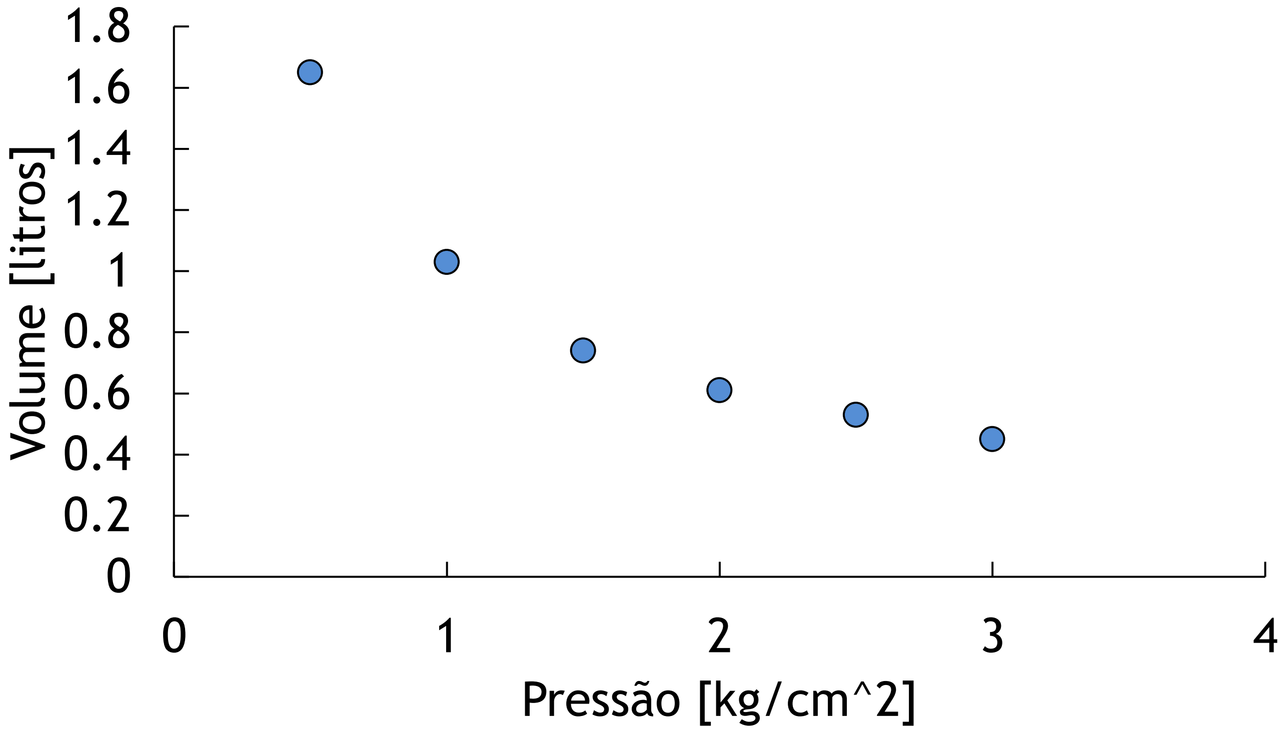
- ¿?



2. Regressão não linear.

- Potencial.

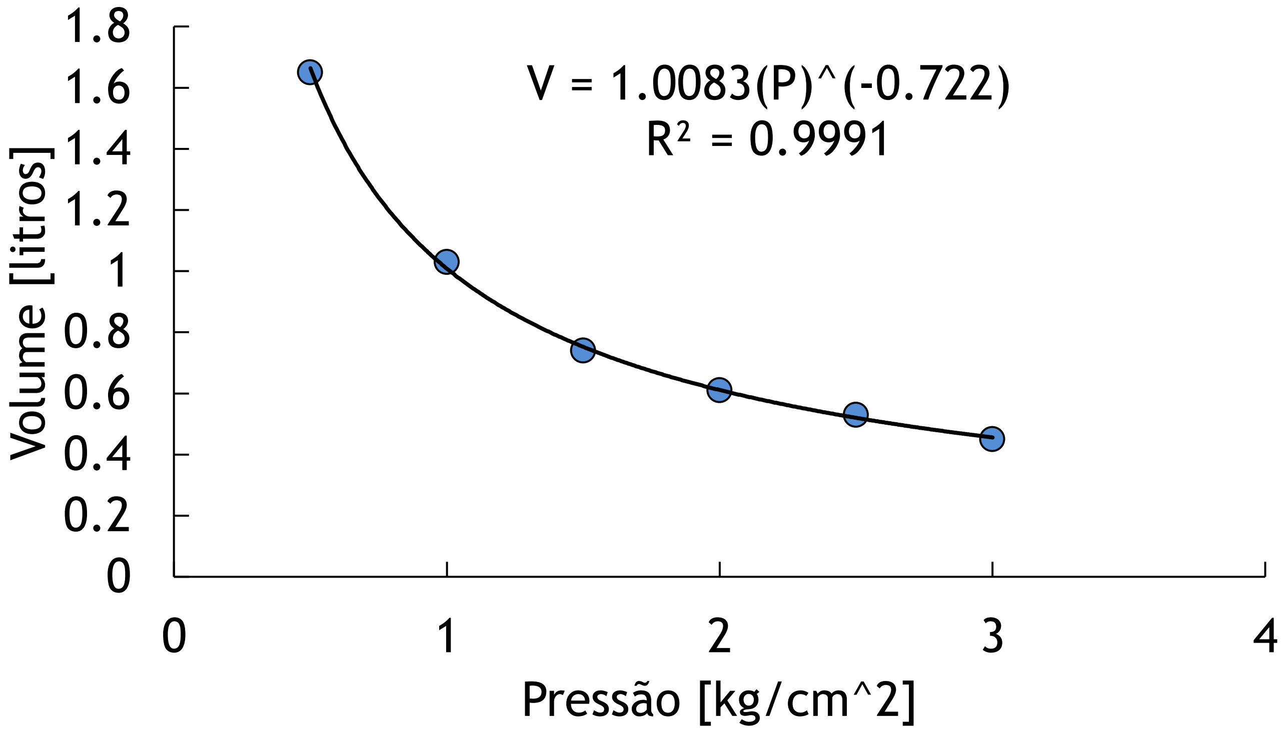
$$y = \alpha(x_1)^{\beta} + \varepsilon$$



2. Regressão não linear.

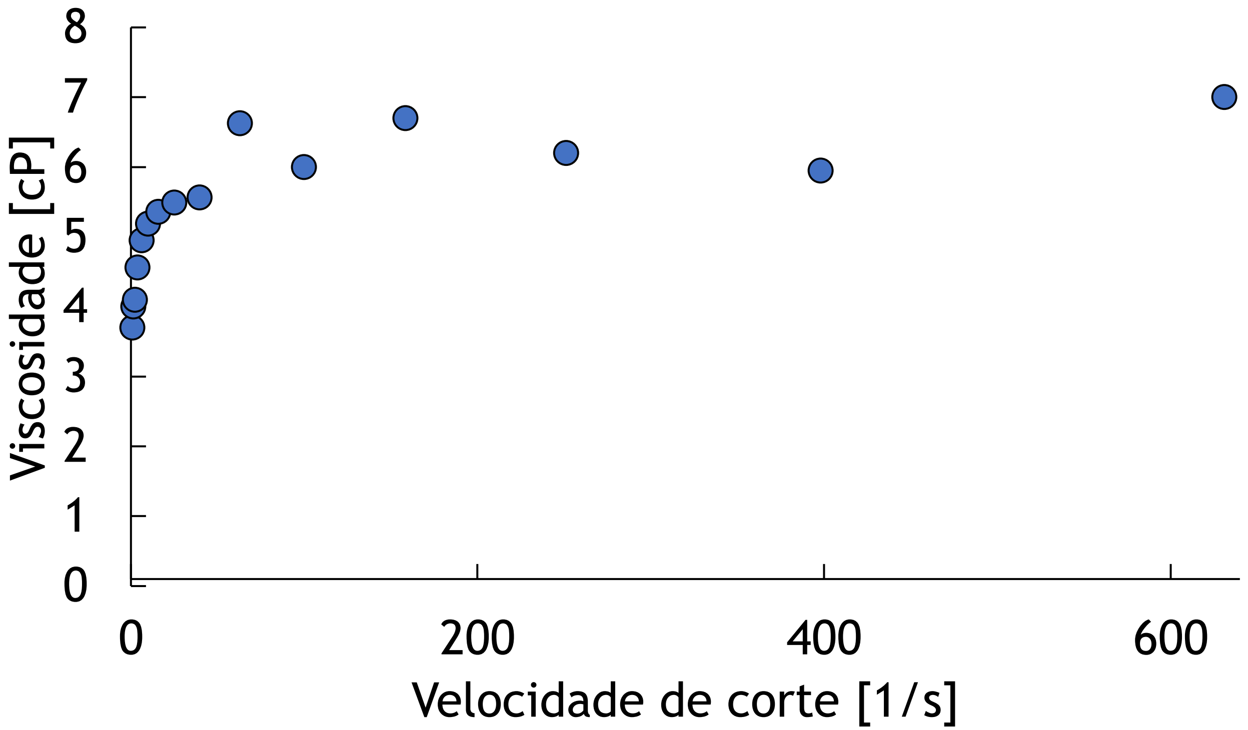
- Potencial.

$$y = \alpha(x_1)^{\beta} + \varepsilon$$



2. Regressão não linear.

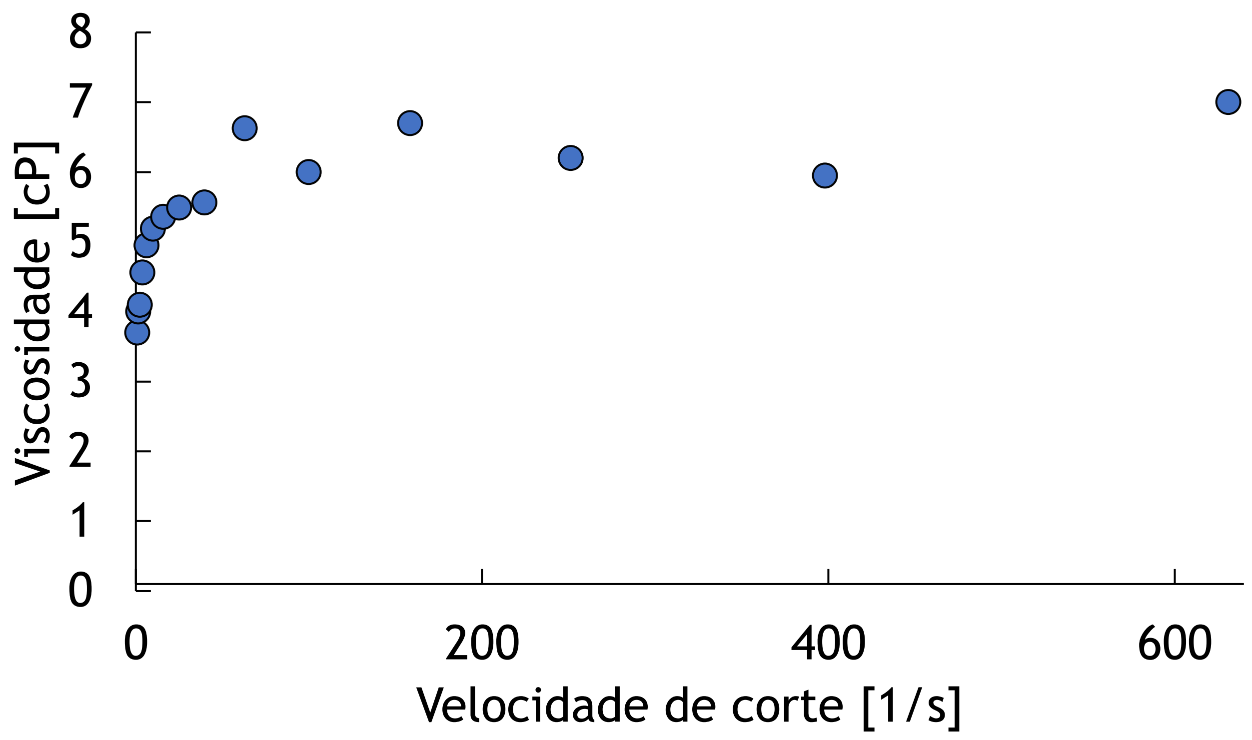
- ¿?



2. Regressão não linear.

- Logarítmica.

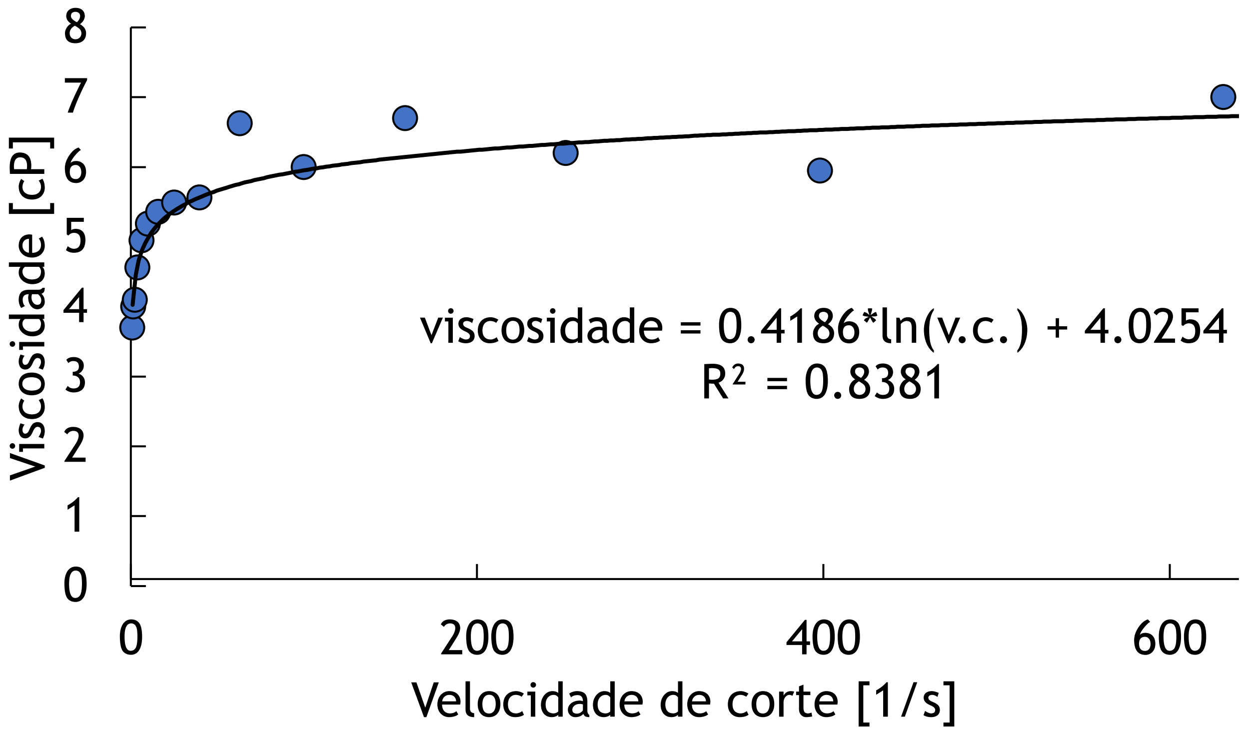
$$y = \alpha \ln(x) + \varepsilon$$



2. Regressão não linear.

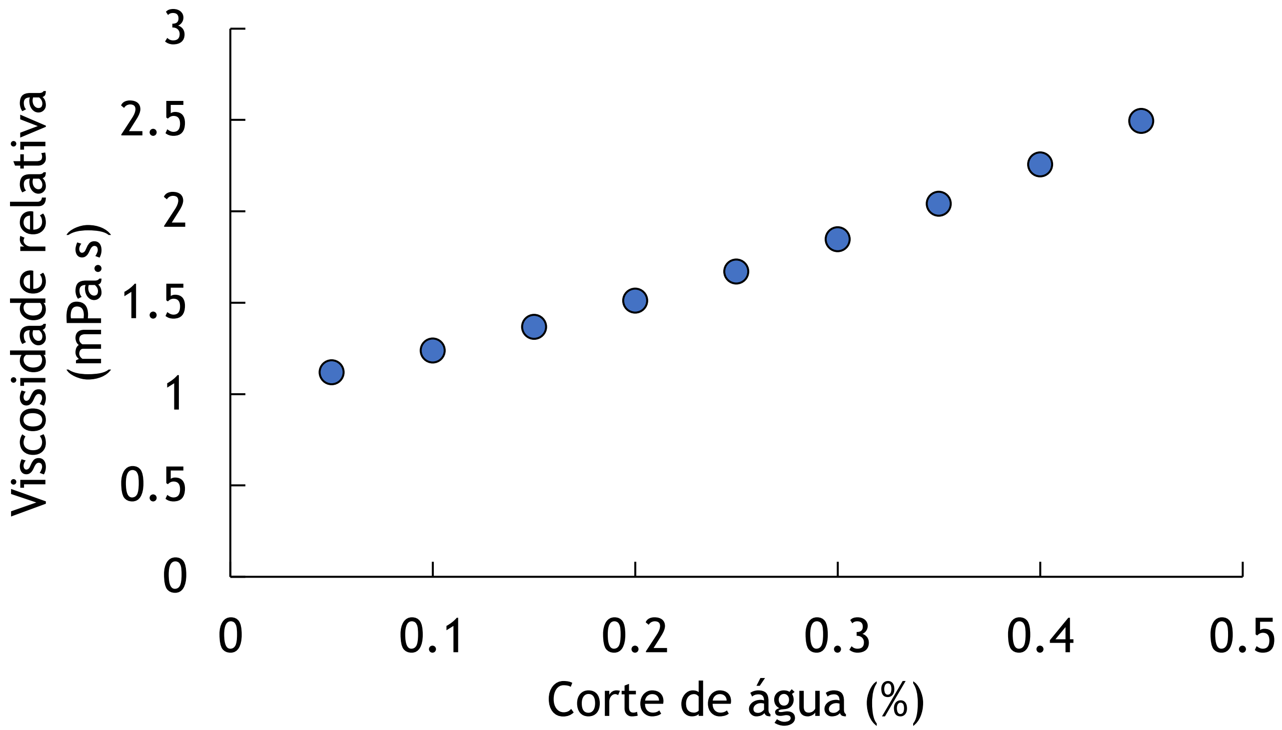
- Logarítmica.

$$y = \alpha \ln(x) + \varepsilon$$



2. Regressão não linear.

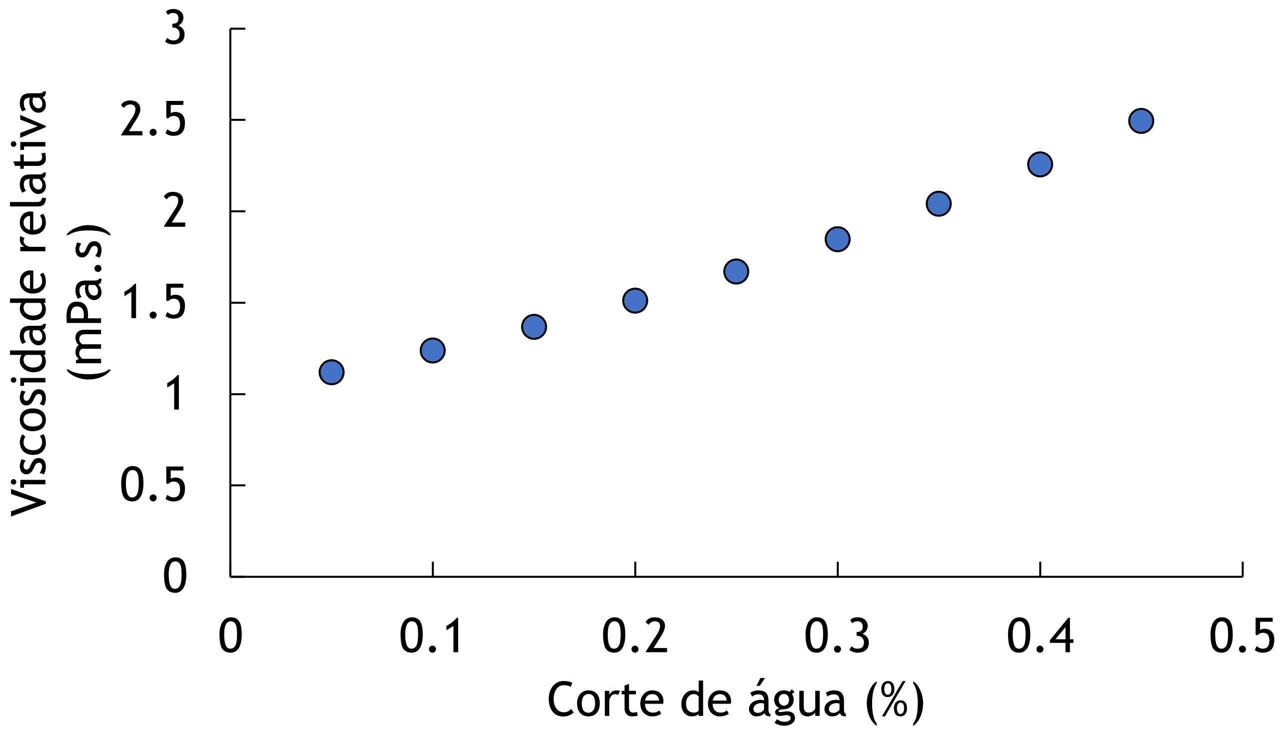
- ¿?



2. Regressão não linear.

- Exponencial

$$\ln(\eta_r) = k_1 + k_2 T + k_3 \phi + k_4 T \phi$$

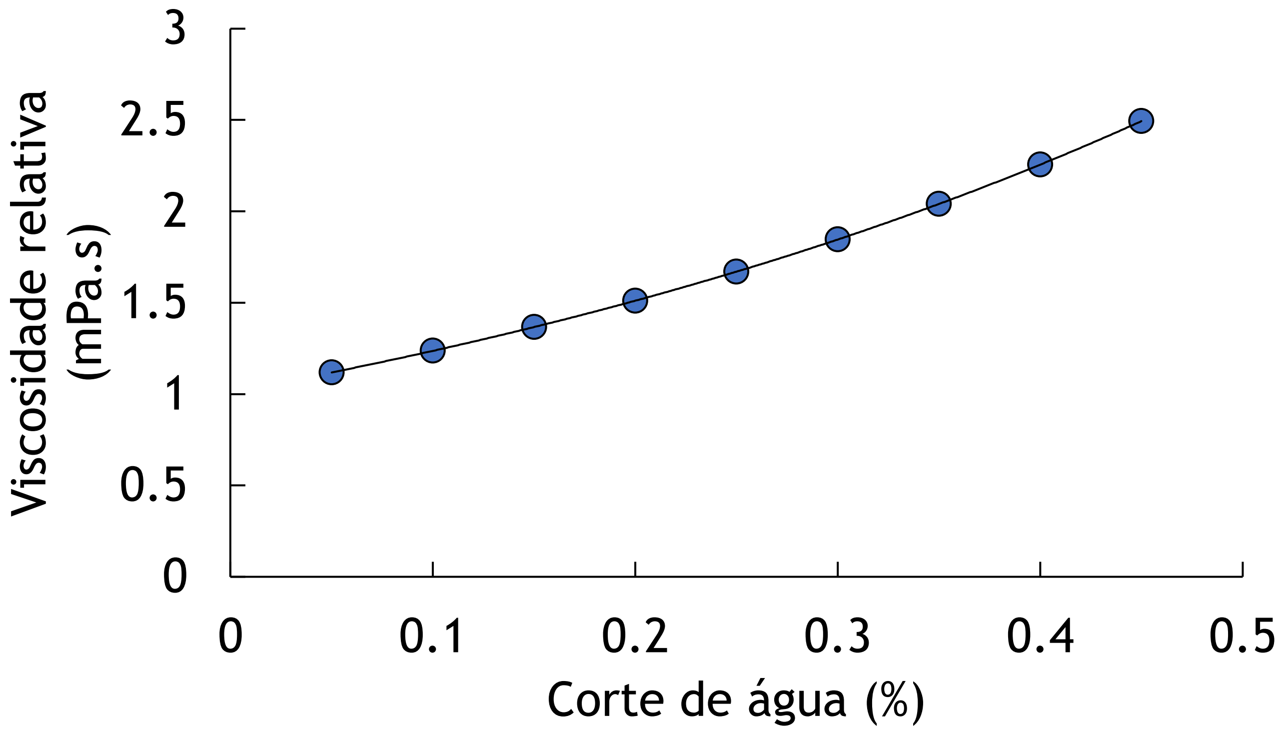


2. Regressão não linear.

- Exponencial

$$\ln(\eta_r) = k_1 + k_2 T + k_3 \phi + k_4 T \phi$$

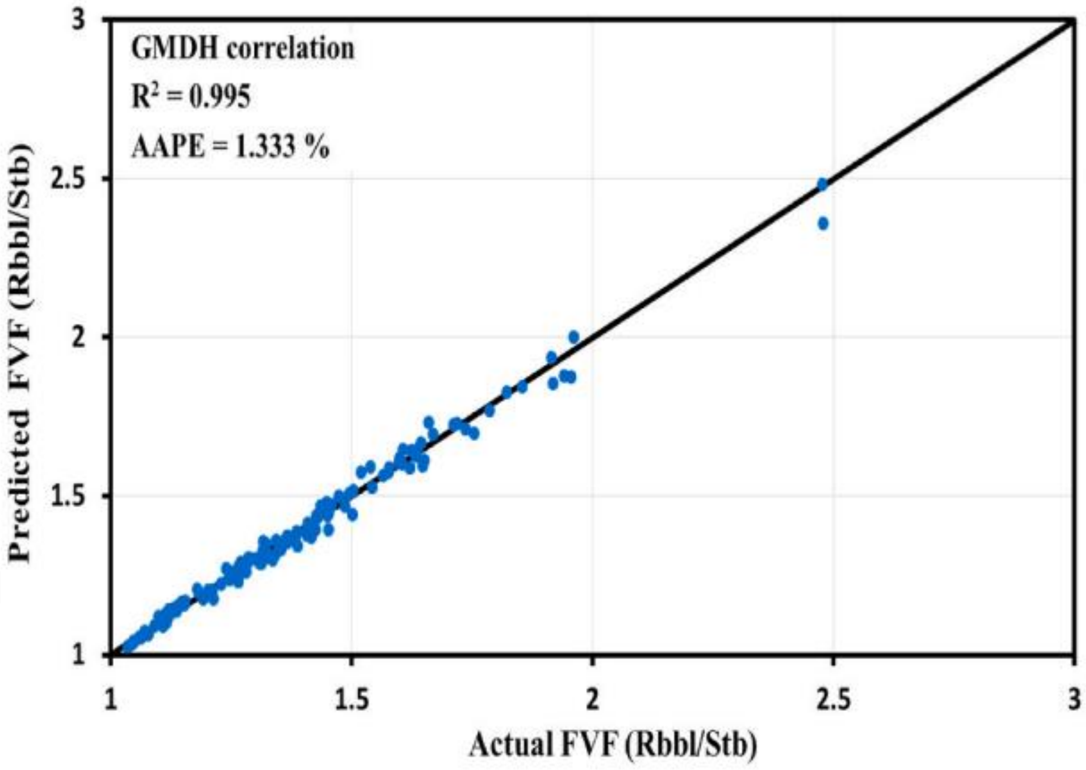
Ronningsen, P. (1995). Correlations for predicting Viscosity of W/O - Emulsions based on North Sea Crude Oils. SPE 28968.



Previsão do fator de volume de petróleo:

$$y = a_0 + a_1 A(x_i) + a_2 B(x_i) + \cdots + a_n D(x_i)$$

$$Rs, T, °API, \gamma_g$$

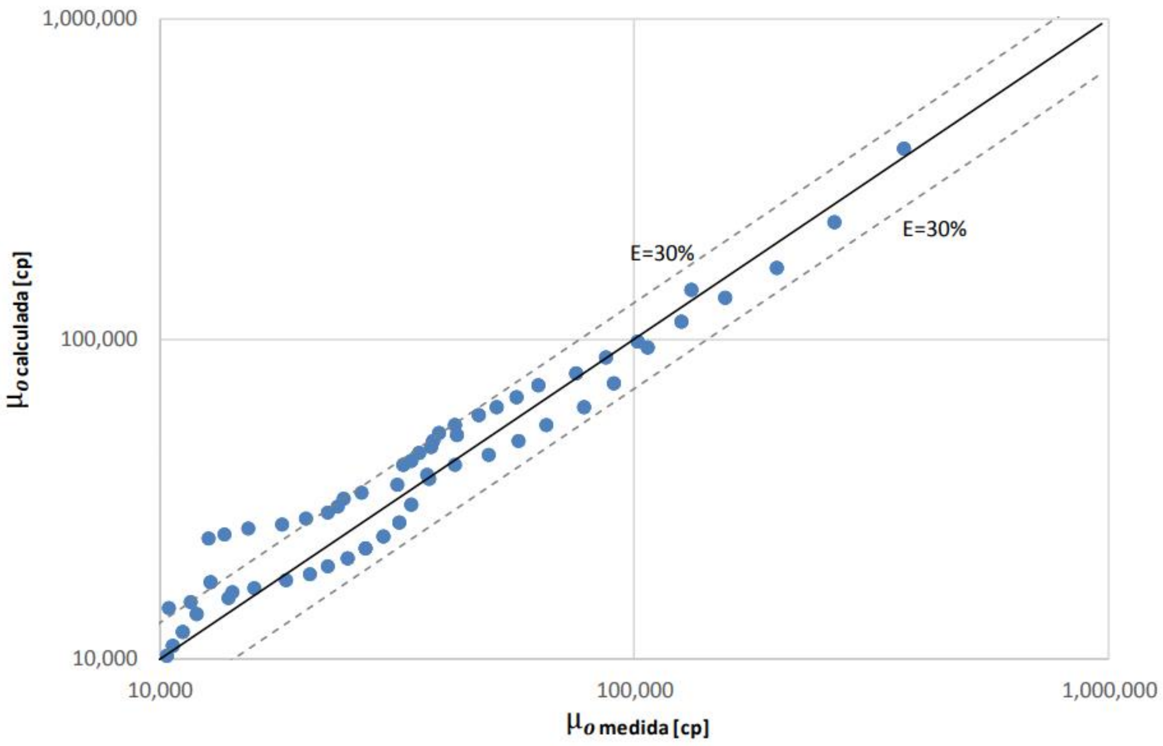


*Ayoub, M. A., Elhadi, A., Fatherlhman, D., Saleh, M. O., Alakbari, F. S., & Mohyaldinn, M. E. (2022). *A new correlation for accurate prediction of oil formation volume factor at the bubble point pressure using Group Method of Data Handling approach. Journal of Petroleum Science and Engineering, 208, 109410.* doi:10.1016/j.petrol.2021.109410

Regressão de López-Hernández (2017):

$$\mu_o = K_1 * e^{K_2 T} \gamma^{(K_3 T^2 + K_4 T + K_5)}$$

$K_{1,2,3}$: constantes de regresión
μ_o: viscosidad del aceite
T: temperatura
γ: velocidad de corte

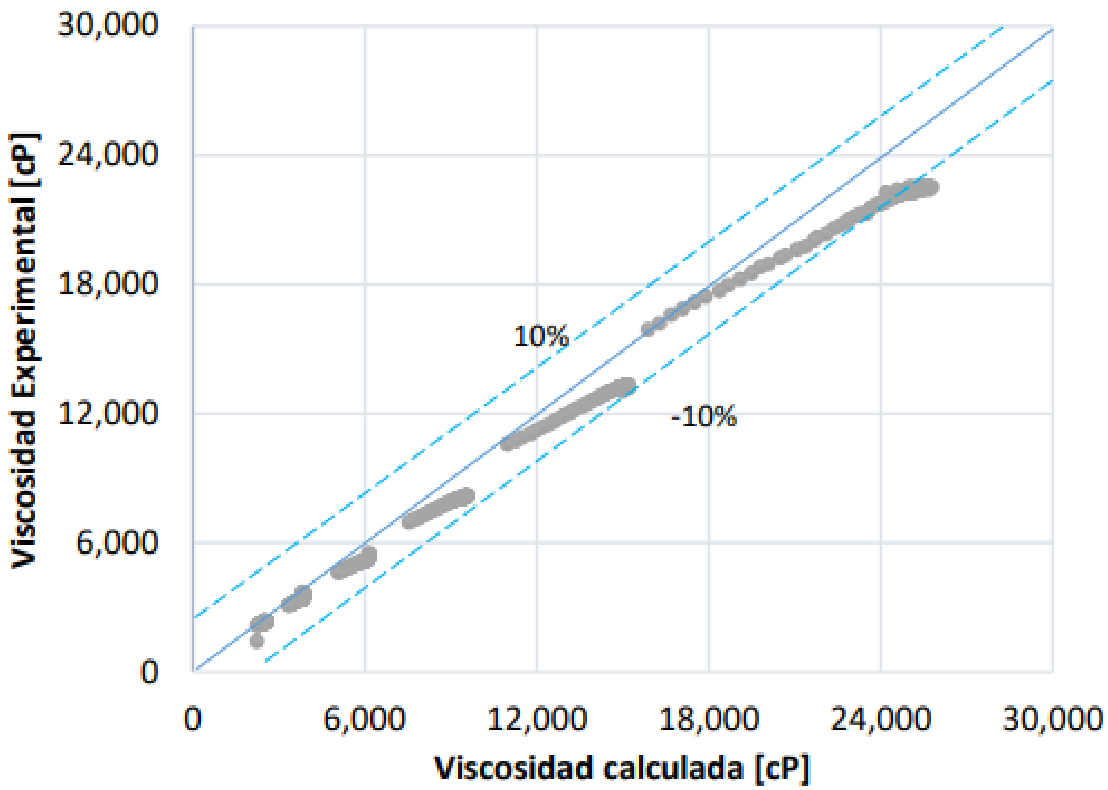


* López-Hernández O. (2017). Estudo Reológico de Emulsões de Água e Petróleo Bruto Pesado de Campos Marinhos no México. Dissertação de mestrado. Universidade Nacional Autônoma do México.

Regressão de López-Pérez (2021):

$$\ln(\mu_r) = K_1 - K_2 \ln(T) + K_3 \emptyset - (K_4 \emptyset \ln(T)\, e^{K_5 \gamma})$$

$K_{1,2,3,4,5}$: constantes de regresión
μ_r: viscosidad del aceite
$\emptyset$: corte de agua
T: temperatura
γ: velocidad de corte



* López-Pérez V. (2021). Obtención de una correlación para determinar la viscosidad de emulsiones agua-aceite en función de temperatura, contenido de agua y velocidad de corte. Tesis de maestría. Instituto Mexicano del Petróleo.

GESTÃO DE DADOS

Visualização:

1. Excel:
   - I. Tradicional.
   - II. Tabelas Dinâmicas.
2. Outras ferramentas:
   - I. OriginLab.
   - II. Python (Matplolib, Seaborn).
   - III. Power BI.
   - IV. Tableu.
   - V. Google Charts.

Formatos diferentes, mas as informações devem ser padronizadas.

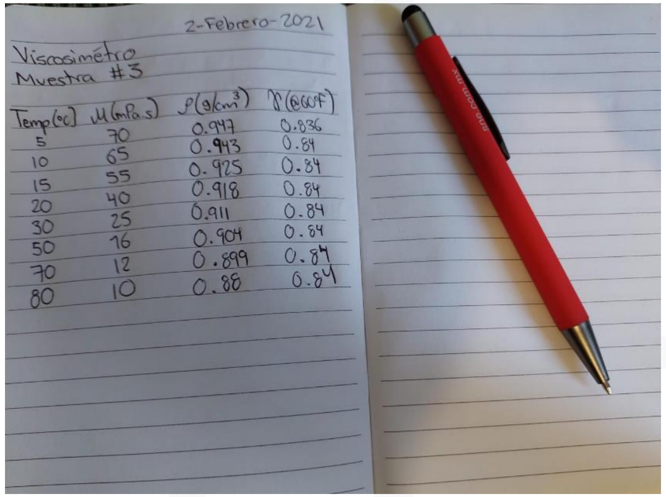
PDF

Data Report

Name	SAMPLE 1 W/W
Operator	Petroleum Analysis Company
Project	Viscosities Study
Instrument	Rheometer
Temperature	35 °C

Temperature °C	Shear Stress Pa	Shear rate 1/s	Viscosity Pa.s
25.00	1.61	2.51	0.0015
25.00	1.88	3.98	0.0025
25.00	2.43	6.31	0.0034
25.01	2.42	10.00	0.0039
25.00	1.95	15.85	0.0043
25.00	2.59	25.12	0.0045
25.00	2.52	39.82	0.0046
24.99	2.07	63.10	0.0047
24.99	2.29	100.01	0.0048
25.00	2.31	158.51	0.0048
25.00	3.73	251.22	0.0049
25.00	7.50	398.16	0.0050
25.00	7.74	631.03	0.0051
25.01	7.77	1000.12	0.0053

Anotações



ASCII

```
NCOLS 6081
NROWS 3801
XLLCENTER 453200
YLLCENTER 4076200
CELLSIZE 5
NODATA_VALUE -999
1658.09 1658.692 1659.284 1659.853 1660.396 1660.917 1661.333
1661.598 1661.816 1662.145 1662.527 1662.989 1663.736 1664.876
1666.333 1667.942 1669.471 1670.768 1671.946 1673.041 1674.209
1675.434 1676.593 1677.707 1678.757 1679.715 1680.629 1681.537
1682.469 1683.468 1684.552 1685.721 1686.971 1688.28 1689.601
1690.948 1692.307 1693.653 1694.957 1696.216 1697.467 1698.642
1699.711 1700.764 1701.815 1702.844 1703.757 1704.509 1705.057
1705.439 1705.711 1705.854 1705.742 1705.445 1705.018 1704.482
1703.776 1702.958 1702.004 1700.902 1699.725 1698.618 1697.54
1696.522 1695.674 1695.12 1694.979 1695.266 1695.841 1696.438
1697.041 1697.646 1698.296 1698.779 1699.122 1699.385 1699.618
```

Abrir arquivos do Excel

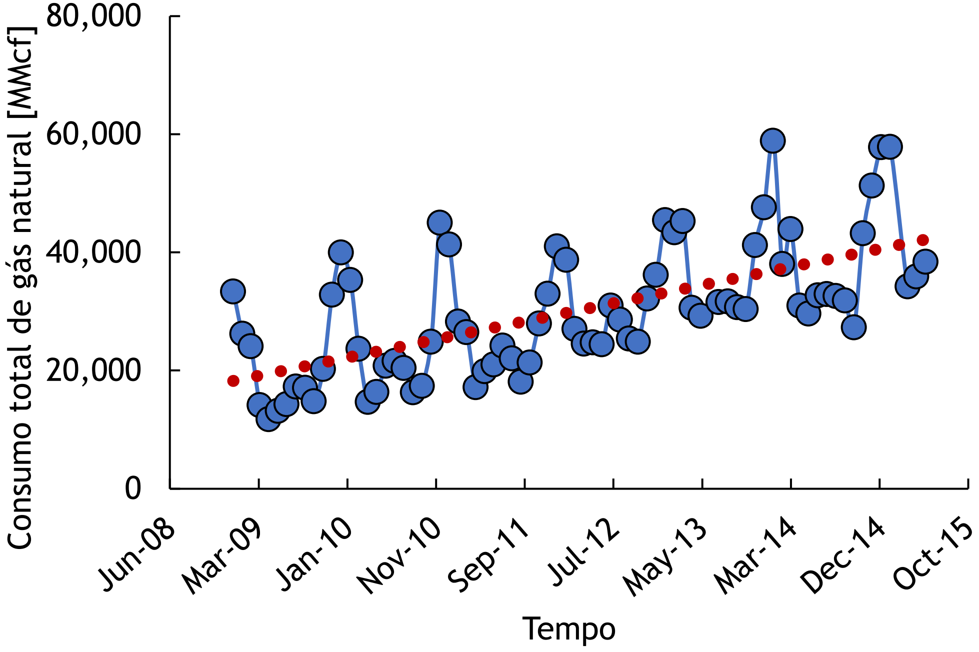
Fontes de dados:

a. Externo:

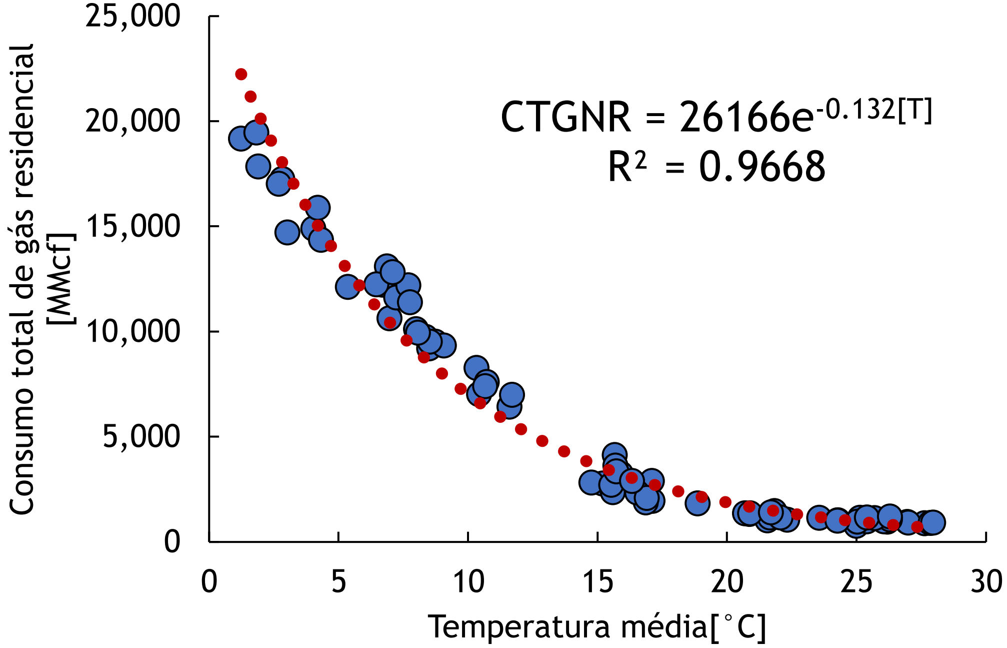
- https://regressit.com

- Consumo de gás natural.
- Temperaturas ambientes mínimas, máximos e médias.

1. Baixar arquivo:
   - NC_natural_gas_consumption_analysis.xlsx
2. Explore os dados.
3. Identificar os dados de interesse.
   - Datas.
   - Consumo mensal de gás natural.
   - Temperaturas máxima, mínima e média.
4. Separe os dados de interesse.

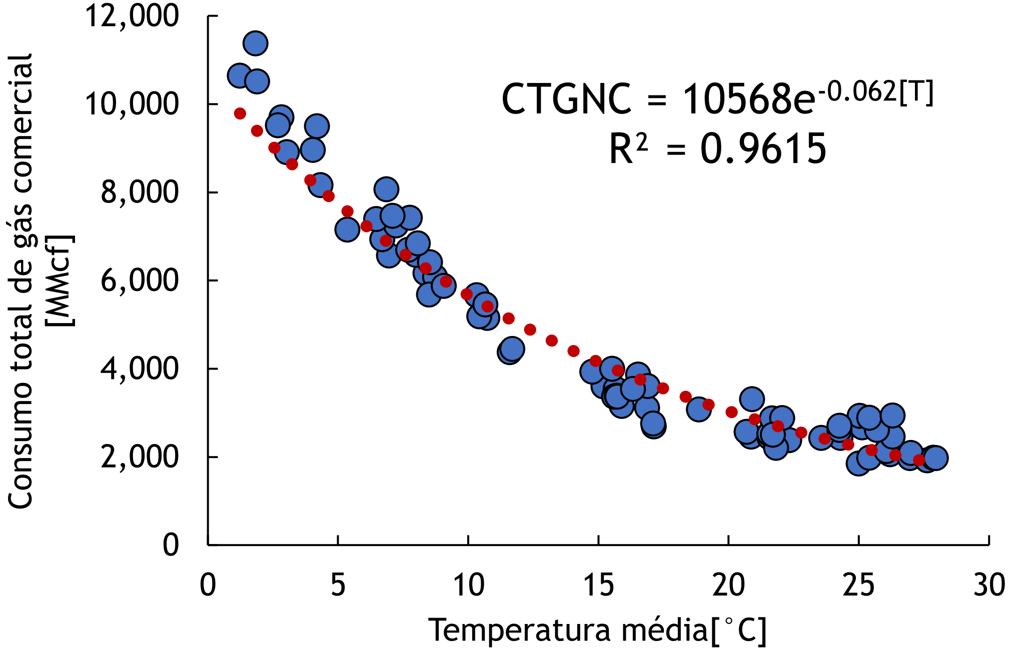


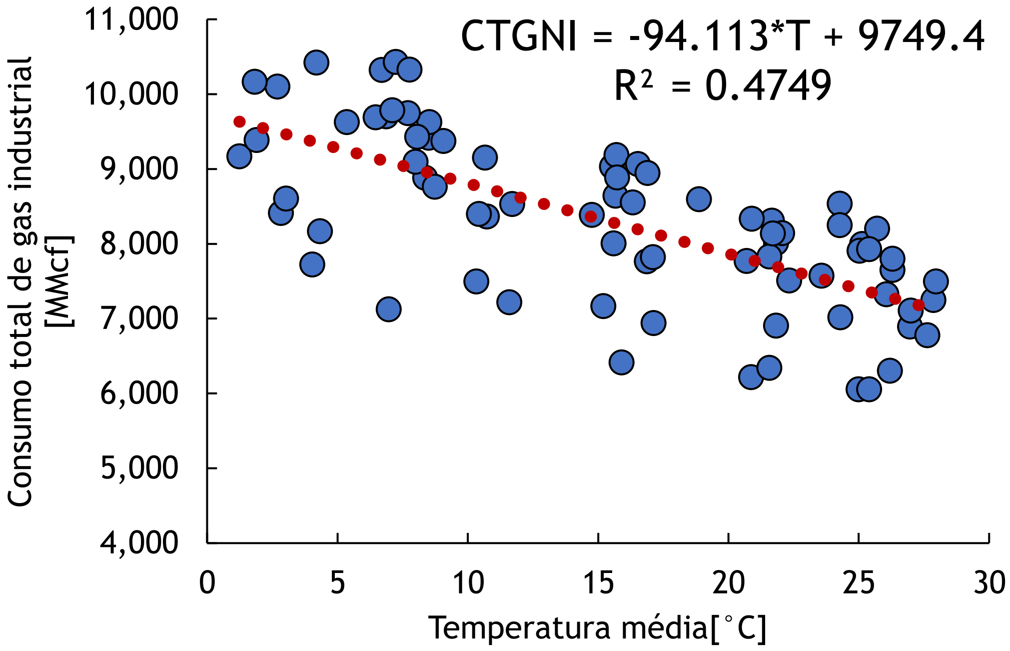
Consumo total de gás natural [MMcf]
80,000
60,000
40,000
20,000
0
Jun-08
Mar-09
Jan-10
Nov-10
Sep-11
Jul-12
May-13
Mar-14
Dec-14
Oct-15
Tempo

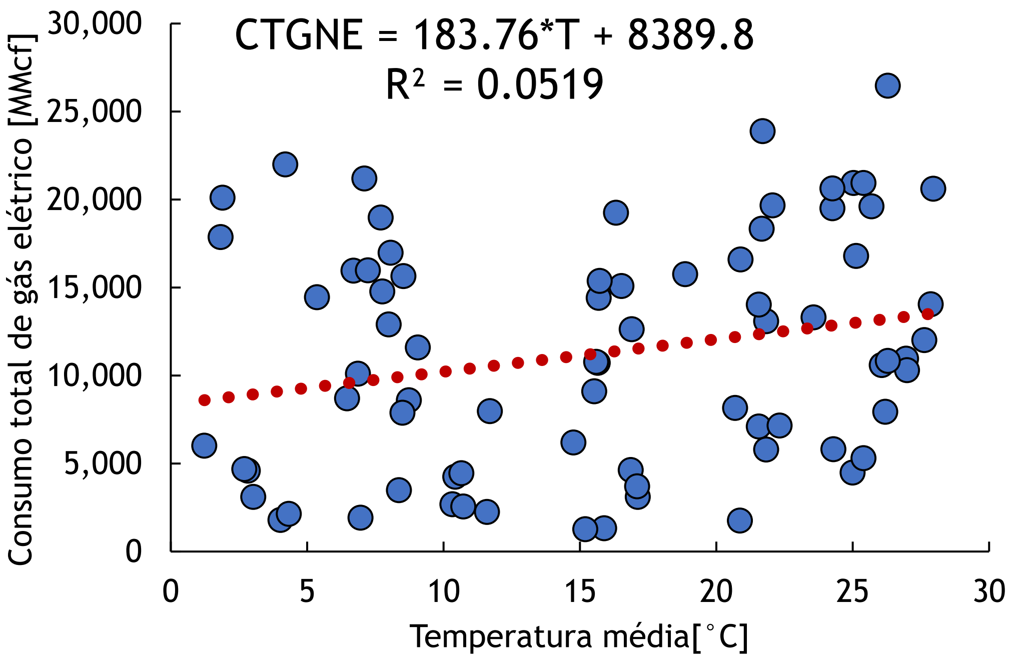


Consumo total de gás residencial [MMcf]
25,000
20,000
15,000
10,000
5,000
0
0
5
10
15
20
25
30
Temperatura média[°C]
CTGNR = 26166e$^{-0.132[T]}$
R^2 = 0.9668

GESTÃO DE DADOS - EXEMPLO



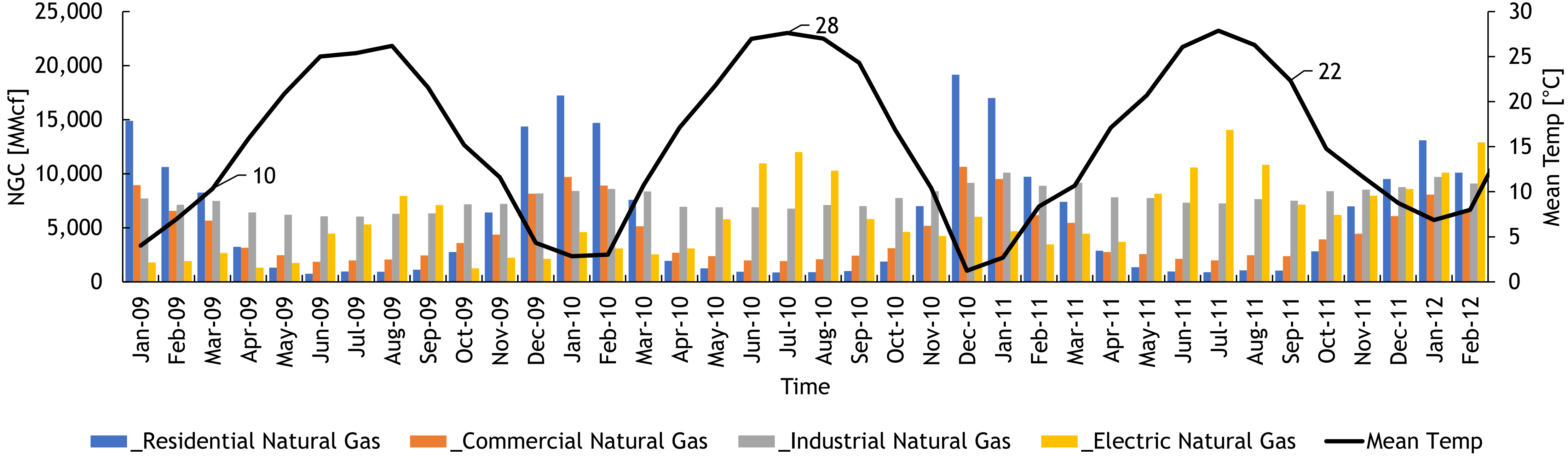




Que modelo global é proposto para o consumo de gás natural em função da temperatura?

GESTÃO DE DADOS - EXEMPLO

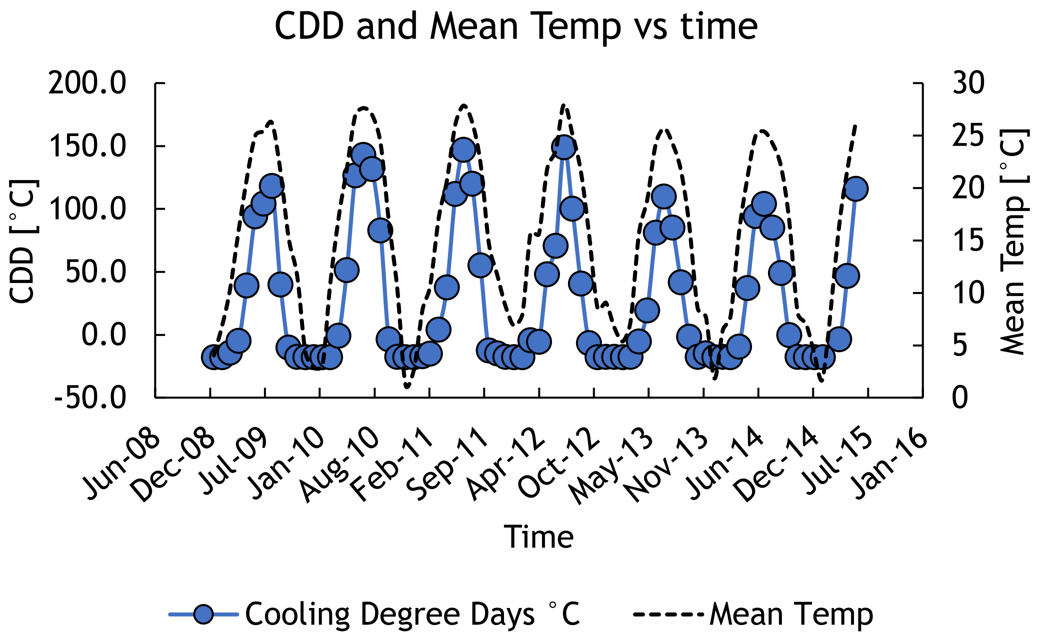
Gráfico de consumo de gás natural por setor e temperatura média

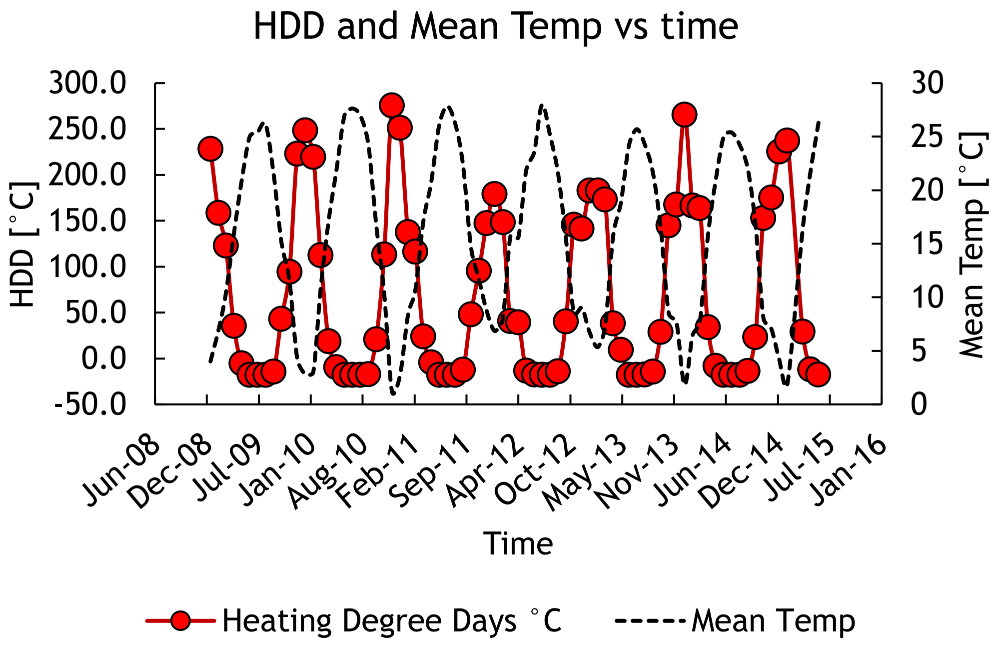


Heating and Cooling Degree Days

Degree days are based on the assumption that when the outside temperature is 65°F, we don't need heating or cooling to be comfortable. Degree days are the difference between the daily temperature mean, (high temperature plus low temperature divided by two) and 65°F. If the temperature mean is above 65°F, we subtract 65 from the mean and the result is Cooling Degree Days. If the temperature mean is below 65°F, we subtract the mean from 65 and the result is Heating Degree Days.

From: https://www.weather.gov/key/climate_heat_cool


CDD and Mean Temp vs time
CDD [˚C]
Mean Temp [˚C]
200.0
150.0
100.0
50.0
0.0
-50.0
30
25
20
15
10
5
0
Jun-08
Dec-08
Jul-09
Jan-10
Aug-10
Feb-11
Sep-11
Apr-12
Oct-12
May-13
Nov-13
Jun-14
Dec-14
Jul-15
Jan-16
Time
Cooling Degree Days ˚C
Mean Temp


HDD and Mean Temp vs time
HDD [˚C]
Mean Temp [˚C]
300.0
250.0
200.0
150.0
100.0
50.0
0.0
-50.0
30
25
20
15
10
5
0
Jun-08
Dec-08
Jul-09
Jan-10
Aug-10
Feb-11
Sep-11
Apr-12
Oct-12
May-13
Nov-13
Jun-14
Dec-14
Jul-15
Jan-16
Time
Heating Degree Days ˚C
Mean Temp

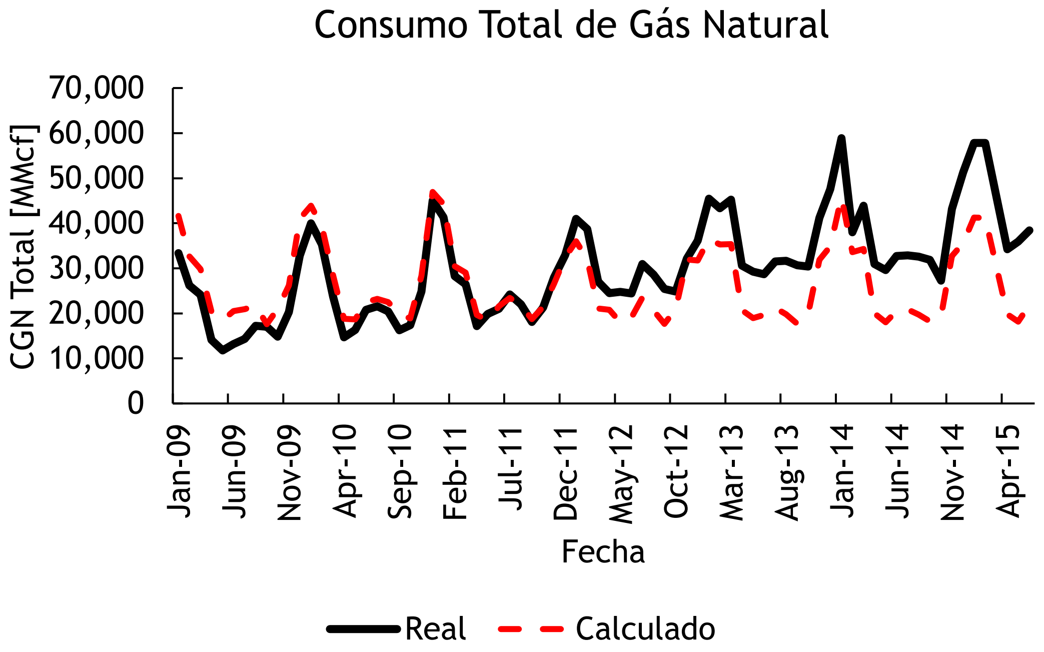
$$CGN_i(CDD, HDD, TM) = ¿?$$

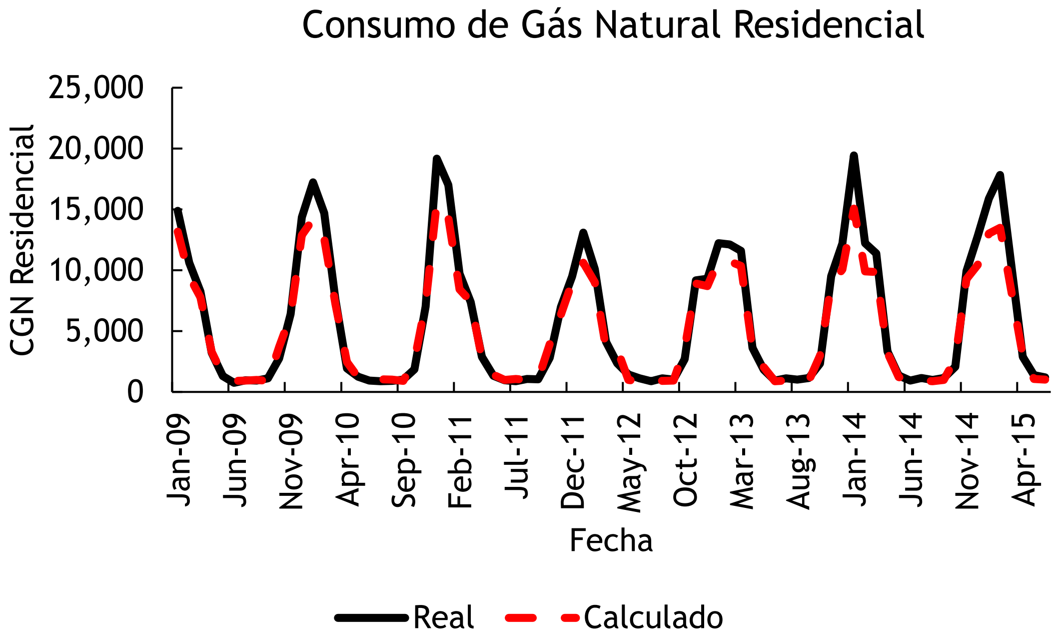
Exemplo de análise dos resultados de Consumo de Gás Natural

$$CGN_i = A(CCD) + B(HDD) + C(MT) + D$$

			X1 = CCD	X2 = HDD	X3 = MT	Ind.	R^2
Regressão	**Var. Ind.**	**Modelo**	**A**	**B**	**C**	**D**	
1	CGN Total	Linear multivariável	6.19	166.50	911.61	50.29	0.534
2	CGN Residencial	Linear multivariável	-1.72	56.06	83.06	4.71	0.986
3	CGN Comercial	Linear multivariável	-8.96	33.85	147.91	8.20	0.964
4	CGN Industrial	Linear multivariável	-28.83	31.55	423.90	23.35	0.476
5	CGN Eléctrico	Linear multivariável	49.41	17.30	120.16	6.49	0.094

Exemplo de análise dos resultados de Consumo de Gás Natural



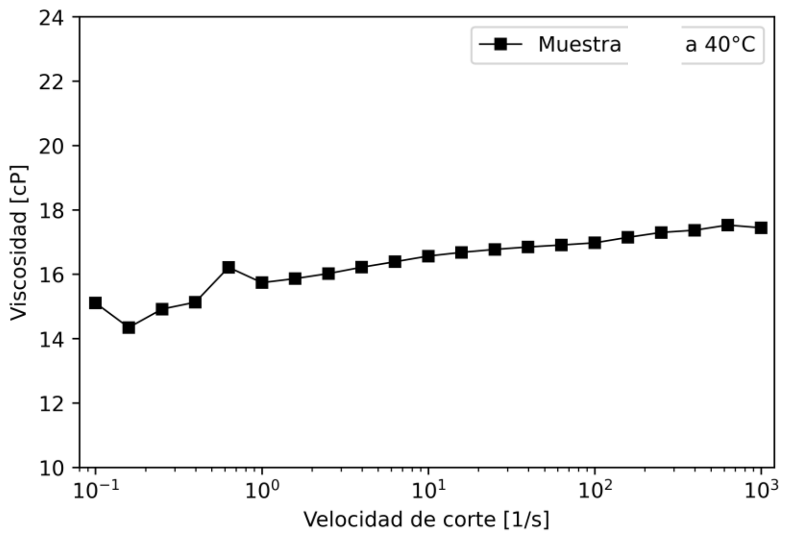


REGRESSÕES

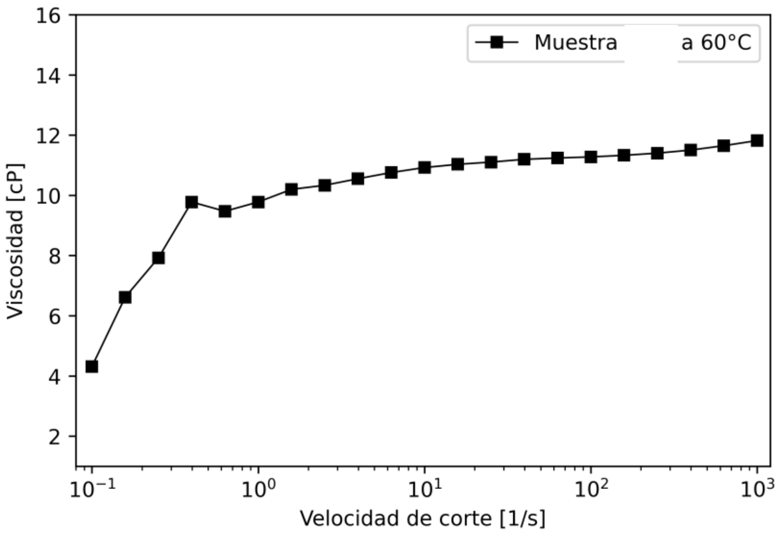
Exemplo #1 de aplicação de regressões em dados de reologia de petróleo

REGRESSÕES - EXEMPLO #1

Amostra K-13, 40°C y 60°C



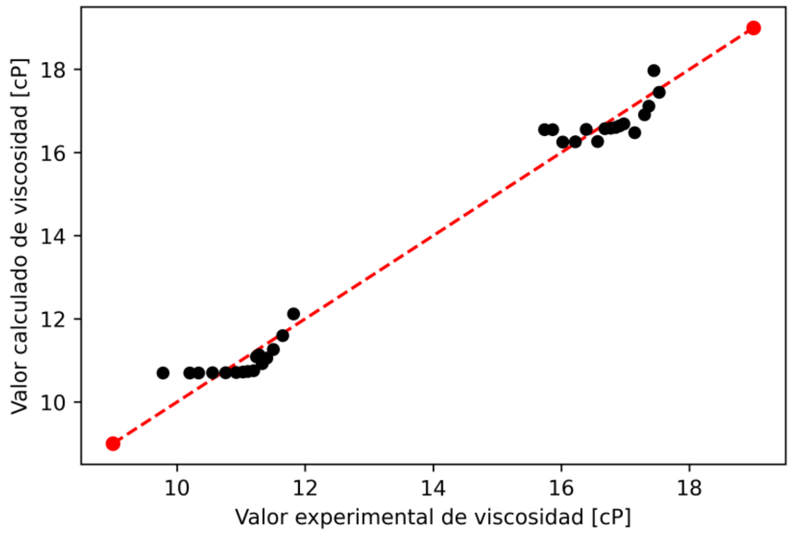
$$\eta(\dot{\gamma}, T, \rho)$$



Velocidade de corte:	[0.1 - 1,000] [1/s]
Viscosidade:	[14.3 - 17.5] [cP]

Velocidade de corte:	[0.1 - 1,000] [1/s]
Viscosidade:	[4.3 – 11.8] [cP]

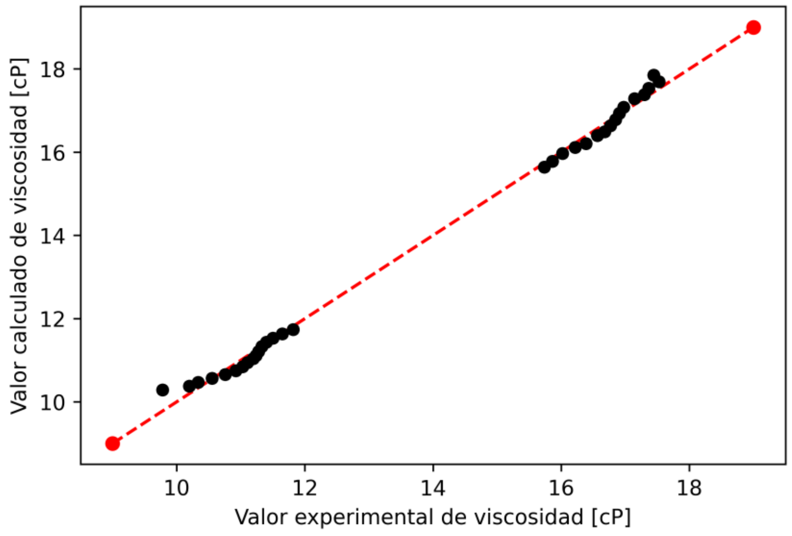
Amostra K-13, 40°C y 60°C



Linear multivariável

$$\eta = A + B(\dot{\gamma}) + C(T) + D(\rho)$$

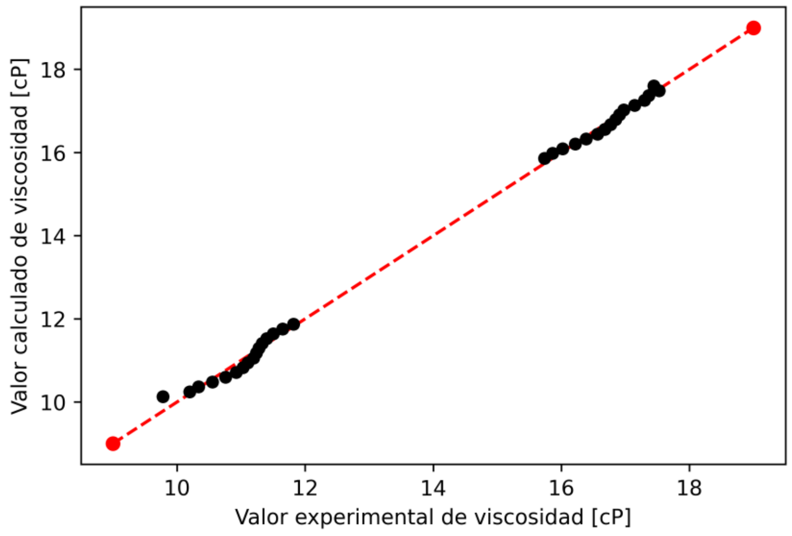
Amostra K-13, 40°C y 60°C



Potencial multivariável

$$\eta = A(\dot{\gamma}^{B})(T^{C})(\rho^{D})$$

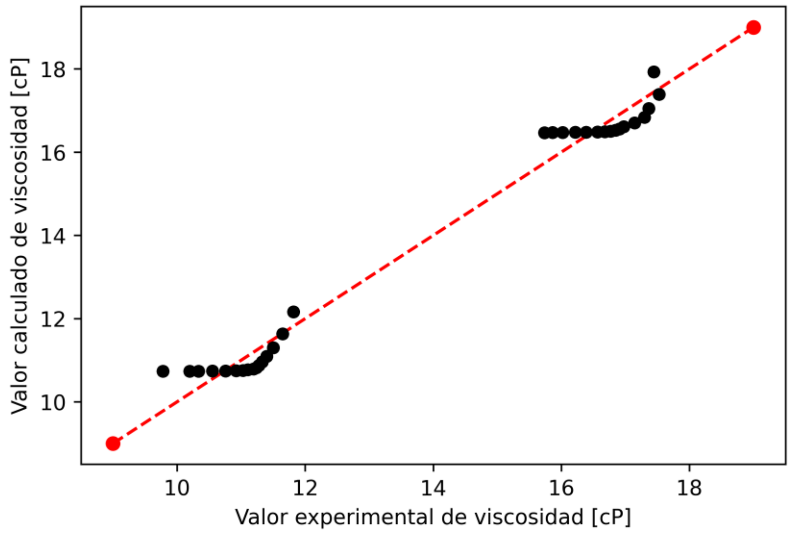
Amostra K-13, 40°C y 60°C



Logarítmico multivariável

$$\eta = A + (B)\log(\dot{\gamma}) + (C)\log(T) + (D)\log(\rho)$$

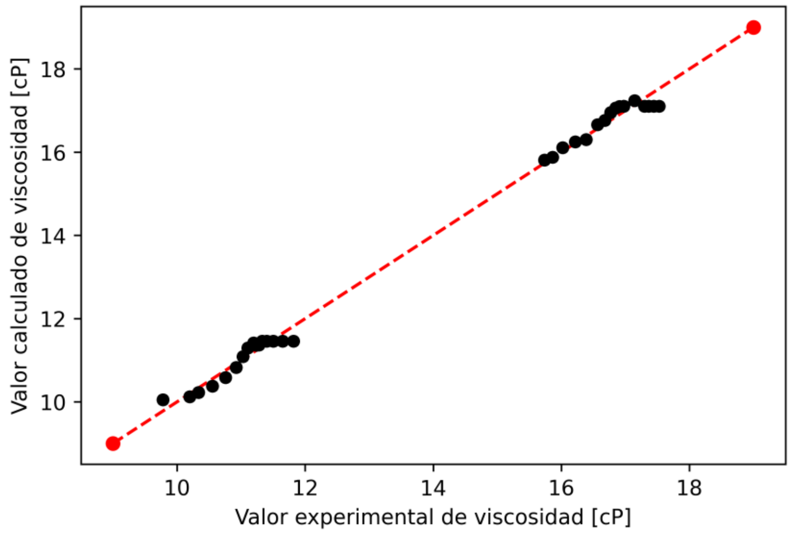
Amostra K-13, 40°C y 60°C



Zhilin et al.(2012)

$$\eta = A + B(\dot{\gamma}) + C(T) + D(\rho)(\dot{\gamma})$$

Amostra K-13, 40°C y 60°C



López-Pérez (2021)

$$\eta = A + B(T) + C(\rho) + D(\rho)\log(T)e^{(E)\dot{\gamma}}$$

Exemplo #2 de aplicação de regressões em dados de reologia de petróleo

Amostra poço	A
Tipo de amostra	Sem centrifugação
Temperatura	15, 40 y 60°C
Velocidade de corte	De 10 a 630 [1/s]
Densidade	De 0.78 a 0.83 [g/cm^3]
Viscosidade	De 1.3 a 5.2 [cP]
Regime de fluxo	Newtoniano

Amostra poço A, 15°C, 40°C y 60°C

Modelo/Autor	Forma funcional
López-Pérez, 2021	$\eta = a + (b)ln(T) + c\rho + \left[(d\rho)\ln(T)\exp^{e\dot{\gamma}}\right]$
Exponencial	$\eta = (a)exp^{(b\dot{\gamma}+cT+d\rho)}$
López-Hernández, 2017	$\eta = (a + bT + c\rho + dT\rho)exp^{e\dot{\gamma}}$
Potencial	$\eta = a(\dot{\gamma}^{b})(T^{c})(\rho^{d})$
Farah et al., 2005	$\eta = a + (b)ln(\dot{\gamma}) + (c)\ln(T) + (d)\ln(\rho)$
Lineal multivariable	$\eta = a + b\dot{\gamma} + cT + d\rho$
Logarítmico	$\eta = (a)ln(\dot{\gamma}) + (b)\ln(T) + (c)\ln(\rho)$
Linearização 3 Modificado	$\eta = aT^{b}\dot{\gamma}^{c}/\rho$
Shigemoto et al., 2006	$\eta = a + bT + (c + d\rho)\dot{\gamma}$
Linearização 2	$1/\eta = (aT\rho)/\dot{\gamma}^{(1-b)}$
Linearização 1	$1/\eta = (T\rho)(a\dot{\gamma} + b)/\dot{\gamma}$
Linearização 3	$1/\eta = (aT\dot{\gamma}^{b})/(\dot{\gamma}\rho)$

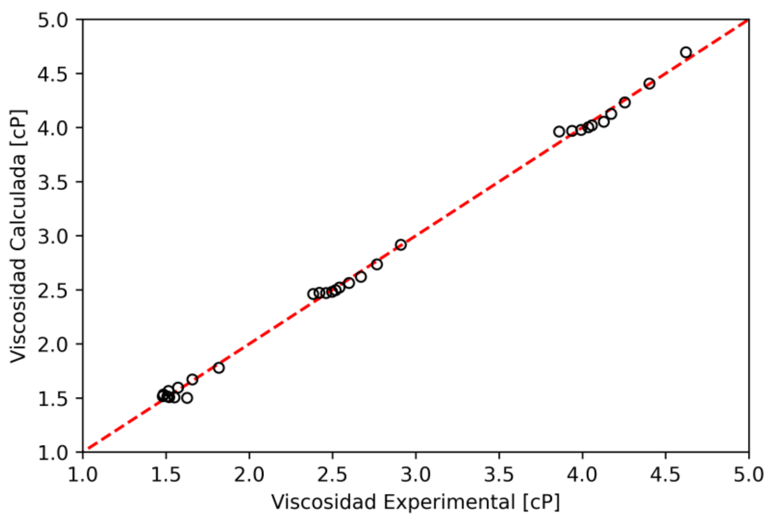
Amostra poço A, 15°C, 40°C y 60°C

Resumo dos parâmetros estatísticos dos modelos

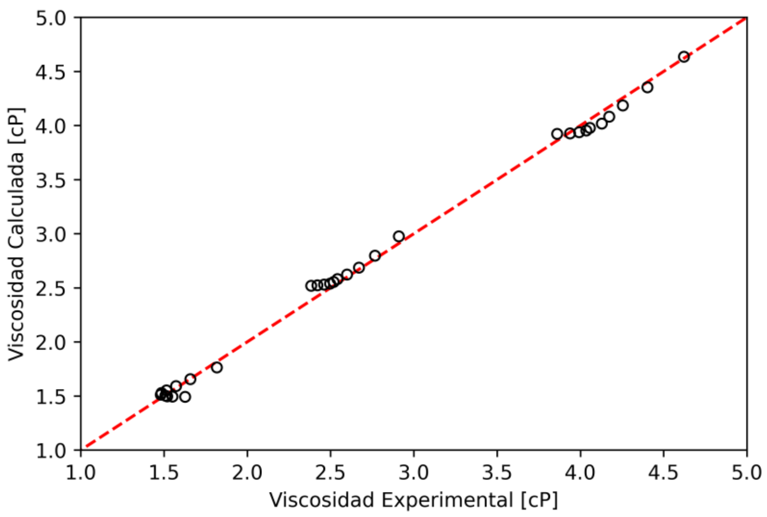
Modelo	R^2	ERP	Menor de 5%	Maior de 5%	Total de dados
1. López-Hernández, 2017	0.99	1.6	96.7%	3.3%	30
2. Exponencial	0.99	2.1	93.3%	6.7%	30
3. Potencial	0.99	2.3	83.3%	16.7%	30
4. López-Pérez, 2021	0.99	2.5	90.0%	10.0%	30

Amostra poço A, 15°C, 40°C y 60°C

1. López-Hernández, 2017



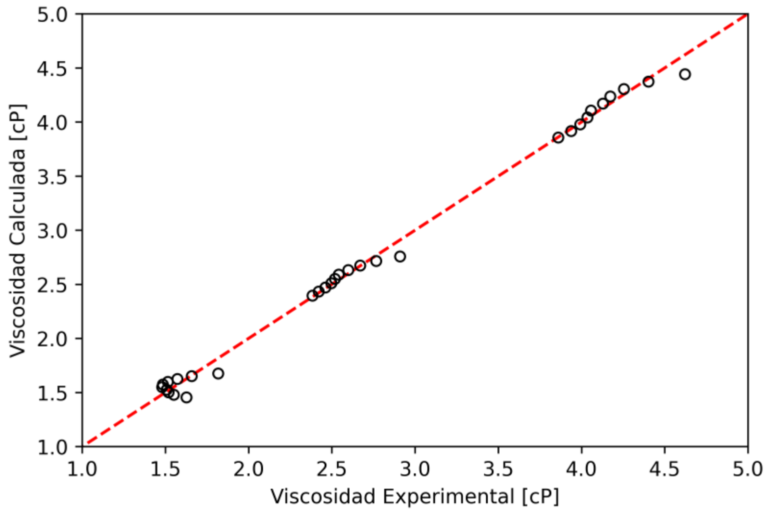
2. Exponencial



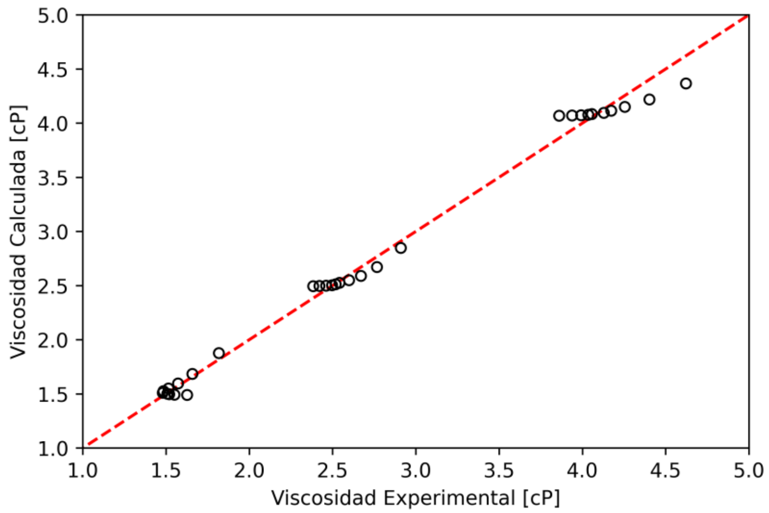
REGRESSÕES - EXEMPLO #2

Amostra poço A, 15°C, 40°C y 60°C

3. Potencial



4. López-Pérez, 2017



Código em Python:

https://classroom.google.com/c/NjI2ODcyMjA1MjE2?cjc=eausxqo

Dados:

https://drive.google.com/drive/folders/185NpRh1fPf_LW3pGsm7W-mIvmoPcqJjC?usp=sharing

É recomendável copiar os códigos para um novo bloco de anotações no Google Colab que pertença ao usuário.

O procedimento é clicar na aba "Arquivo" e selecionar "Novo Caderno".

Estrutura do código

Abaixo estão as seções do código, bem como as partes que precisam ser modificadas de acordo com a configuração do problema que precisa ser resolvido.

Bibliotecas são um conjunto de arquivos (códigos) que são usados para facilitar o desenvolvimento de software.
Eles fornecem funcionalidades e módulos que facilitam a solução de problemas padrão.

```
import numpy as np
from scipy.optimize import minimize
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Esta seção não requer modificação.

2. CARREGANDO DADOS DE ENTRADA

Quando o banco de dados de entrada estiver pronto, os dados serão carregados no programa.
Arquivo no formato "CSV" é necessário.

```
data = '/content/oil_data.csv'
df = pd.read_csv(data)
```

Esta seção requer modificações:
Coloque o local do arquivo de entrada.

Esta seção requer modificações:
Coloque os nomes dos cabeçalhos de coluna do arquivo de entrada em cada uma das matrizes definidas no programa.

```
xm1 = np.array(df["WTI_PRICE"])
xm2 = np.array(df["HH_PRICE"])
xm3 = np.array(df["NGL_PRICE"])

ym = np.array(df["BEST_PRICE"])
```

	A	B	C	D
1	WTI_PRICE	HH_PRICE	NGL_PRICE	BEST_PRICE
2	27.24	2.4	0.555	26.602
3	29.21	2.66	0.596	28.8013
4	29.92	2.78	0.512	29.0381
5	25.78	3.04	0.469	25.0476
6	28.78	3.59	0.512	28.3888
7	31.86	4.29	0.555	31.1513
8	29.97	3.99	0.551	29.686
9	31.31	4.43	0.583	30.7563

O nome e o número de coeficientes a serem utilizados são declarados. O modelo também é declarado.

Esta seção requer modificações:

O modelo no qual a aproximação da regressão é necessária é escrito.

```
def calc_y(x):
    a = x[0]    # Coeficiente a
    b = x[1]    # Coeficiente b
    c = x[2]    # Coeficiente c
    d = x[3]    # Coeficiente d

    y = a * ( xm1**b ) * ( xm2 ** c ) * ( xm3 ** d )

    return y
```

$$y = A(x_1^b)(x_2^c)(x_3^d)$$

É realizada a definição da função a ser otimizada, bem como os valores iniciais dos coeficientes da regressão.

Esta seção não requer modificação.

```
def objective(x):
    # calcular y
    y = calc_y(x)
    # calcular objective
    obj = 0.0
    for i in range(len(ym)):
        obj = obj + ((y[i]-ym[i])/ym[i])**2
    # return result
    return obj
```

O domínio dos coeficientes de regressão é definido.

O solucionador e a solução da regressão são definidos.

Esta seção requer modificação nos limites dos coeficientes se os resultados exigirem (tentativa e erro).

```
# Suposiciones iniciales
x0 = np.zeros(4)
x0[0] = 0.0 # a
x0[1] = 0.0 # b
x0[2] = 0.0 # c
x0[3] = 0.0 # d
```

```
my_bnds = (-100.0, 100.0)

bnds = (my_bnds, my_bnds, my_bnds, my_bnds)
# Programación secuencial de mínimos cuadrados SLSQP
solution = minimize(objective, x0, method='SLSQP', bounds=bnds)

solution = minimize(objective,x0)

x = solution.x
y = calc_y(x)
```

7. IMPRESSÃO DOS COEFICIENTES DE REGRESSÃO E R^2

Os resultados da solução e do R^2 são impressos.

Esta seção não requer modificação.

```
print('Solucion')

cA = 'A = ' + str(x[0])
print(cA)
cB = 'B = ' + str(x[1])
print(cB)
cC = 'C = ' + str(x[2])
print(cC)
cD = 'D = ' + str(x[3])
print(cD)

from scipy import stats
slope, intercept, r_value, p_value, std_err = stats.linregress(ym,y)
r2 = r_value**2
cR2 = "R^2 de correlacion = " + str(r_value**2)
print(cR2)
```

Os valores obtidos vs os valores calculados são grafados, assim como a linha de referência.

Esta seção requer modificação.
O domínio da linha de referência é modificado para que o intervalo completo de resultados possa ser exibido.

```
plt.figure(1)

plt.plot([10,140],[10,140],'k-',label='Measured')

plt.title('Valor real (YM) vs Predecido (Y)')
plt.plot(ym,y,'o')
plt.xlabel('Valor medido (YM)')
plt.ylabel('Valor predecido (Y)')
plt.grid(True)
plt.show()
```

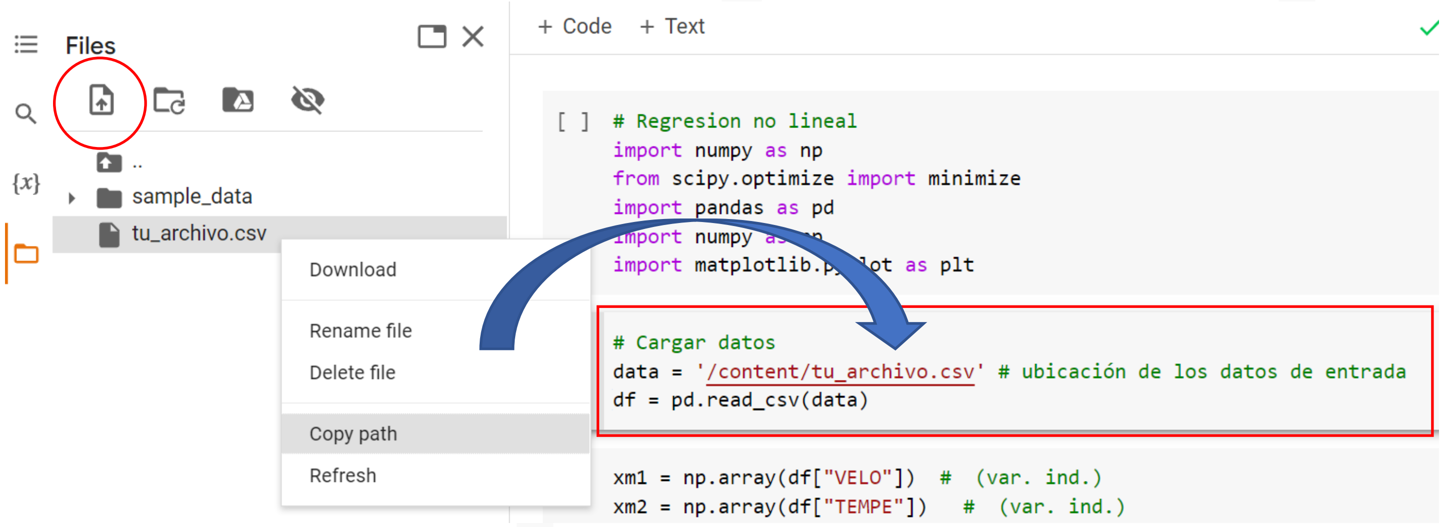
1. Variável alvo: viscosidade [cP].
2. Variáveis independentes: velocidade de corte [1/s], temperatura [°C] e densidade [g/cm^3].
3. O conjunto de dados de entrada é uma amostra de 10 amostras de óleo diferentes.
4. As temperaturas dos dados são de 25, 50 e 80°C.
5. As velocidades de corte variam de 2,51 a 1.000 [1/s].
6. As densidades variam de 0,77 a 0,92 [g/cm^3].
7. Um modelo exponencial da forma é proposto:

$$\eta = A(\dot{\gamma}^{B})(T^{C})(\rho^{D})$$

8. Preparar dados de entrada em uma planilha do Excel (extrair dados de "3_Viscosidad_Base.csv")
9. Lembre-se de que as primeiras colunas do arquivo devem ser as variáveis independentes.
10. Lembre-se de que a última coluna deve corresponder à variável dependente.
11. Salve o arquivo no formato: CSV (separado por vírgulas).

	V.Ind.			V. Dep.	
	A	B	C	D	E
1	VELO	TEMPE	DENSI	VISCO	
2	2.51219	25	0.909	13.46228	
3	3.98155	25	0.909	15.54542	
4	6.31033	25	0.909	16.40941	
5	10.0012	25.01	0.909	16.91292	
6	15.8508	25	0.909	17.2562	

12. Ir para o Google Colab (arquivo "3_Ejemplo_4_Visco_Todas_Temp.ipynb")
13. Carregue o arquivo CSV da etapa anterior.
14. Copie o caminho do arquivo (/content/tu_archivo.csv).
15. Cole o caminho no bloco de código de carga de dados.



6. Após a seção de carregamento de dados no código, verifique se os cabeçalhos gravados na instrução de matriz correspondem aos gravados no arquivo CSV que foi carregado.
7. Insira o modelo para o qual você deseja executar a regressão.

```
xm1 = np.array(df["VELO"])  #  (var. ind.)
xm2 = np.array(df["TEMPE"])   #  (var. ind.)
xm3 = np.array(df["DENSI"])          #  (var. ind.)

ym = np.array(df["VISCO"])  # (var. objetivo)

# Calculando modelo propuesto
def calc_y(x):
    a = x[0]
    b = x[1]
    c = x[2]
    d = x[3]

    y = a * ( xm1 ** b ) * ( xm2 ** c ) * ( xm3 ** d ) # Regresión potencial
    return y
```

M7

	A	B	C	D	E
1	VELO	TEMPE	DENSI	VISCO	
2	2.51219	25	0.909	13.46228	
3	3.98155	25	0.909	15.54542	
4	6.31033	25	0.909	16.40941	
5	10.0012	25.01	0.909	16.91292	
6	15.8508	25	0.909	17.2562	

Modelo

8. Execute o programa clicando no símbolo "play" em todos os blocos de código de forma decrescente.

```
# Regresion no lineal
import numpy as np
from scipy.optimize import minimize
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[3]  # Cargar datos
     data = '/content/6_Ejemplo_4_Visco_Todas_Temp.xlsx.csv' # ubicación de los datos de entrada
     df = pd.read_csv(data)
```

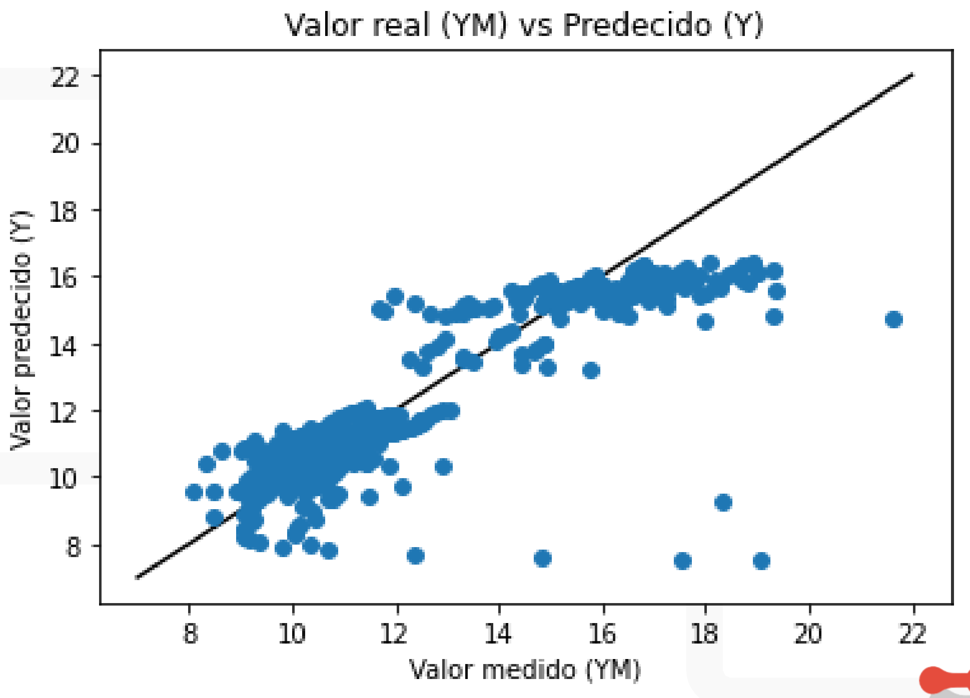
```
xm1 = np.array(df["VELO"])  #  (var. ind.)
xm2 = np.array(df["TEMPE"])   #  (var. ind.)
xm3 = np.array(df["DENSI"])          #  (var. ind.)

ym = np.array(df["VISCO"])  # (var. objetivo)

# Calculando modelo propuesto
```

9. Ao final do último bloco de código é possível visualizar os resultados: coeficientes de regressão, R^2 e gráfico da variável calculada vs variável real.

```
Solucion
A = 19.074265749320542
B = 0.006369106592363837
C = -0.002168512964455647
D = 42.131902949314124
R^2 de correlacion = 0.7778296785295723
```



ANÁLISE DOS RESULTADOS

Critérios de avaliação de regressão:

1. Calcular os valores-alvo com os coeficientes de regressão e com o modelo proposto.
2. Plote os valores calculados versus os valores reais e a linha de referência.
3. Calcule o erro relativo médio (ERP).
4. Calcule o desvio médio (DP).
5. Se necessário, avalie uma amostra dos dados para comparar os gráficos dos dados reais versus os dados estimados com base em qualquer uma das variáveis independentes.

Critérios de avaliação de regressão:

1. Calcular os valores-alvo com os coeficientes de regressão e com o modelo proposto.

Modelo

$$y = A(x_1^B)(x_2^C)(x_3^D)$$

Coeficientes de regresión

A	B	C	D
69.3	0.09	-0.56	16.4

Variáveis independentes

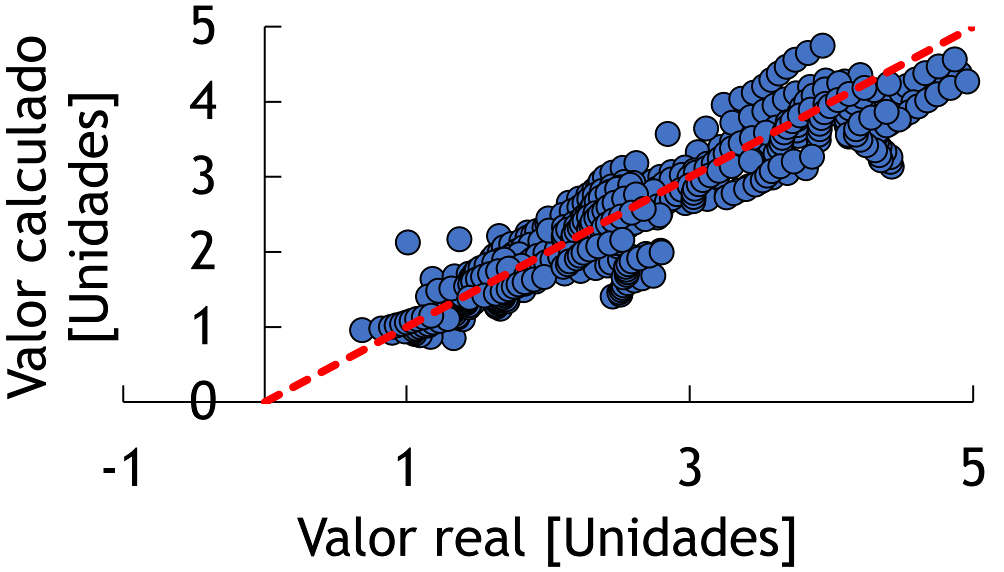
x_1	x_2	x_3
11.00	25.10	0.832
16.85	25.04	0.832
26.12	25.00	0.832
38.81	24.99	0.832
…	…	…

Variável calculada

y
5.72
5.79
5.87
5.95
…

Critérios de avaliação de regressão:

2. Plotar valores calculados vs valores reais e linha de referência.



Critérios de avaliação de regressão:

3. Calcular o erro relativo médio(ERP):

   I. Calcular o erro relativo de cada dado calculado em relação aos dados reais:

$$ER[\%] = \left|\frac{Calculado - Real}{Real}\right| * 100$$

   II. Calcular o erro relativo médio:

$$ERP[\%] = \frac{\sum ER}{Número\ de\ datos}$$

Critérios de avaliação de regressão:

4. Calcular o desvio médio(DP):

   I. Calcular o desvio de cada dado:

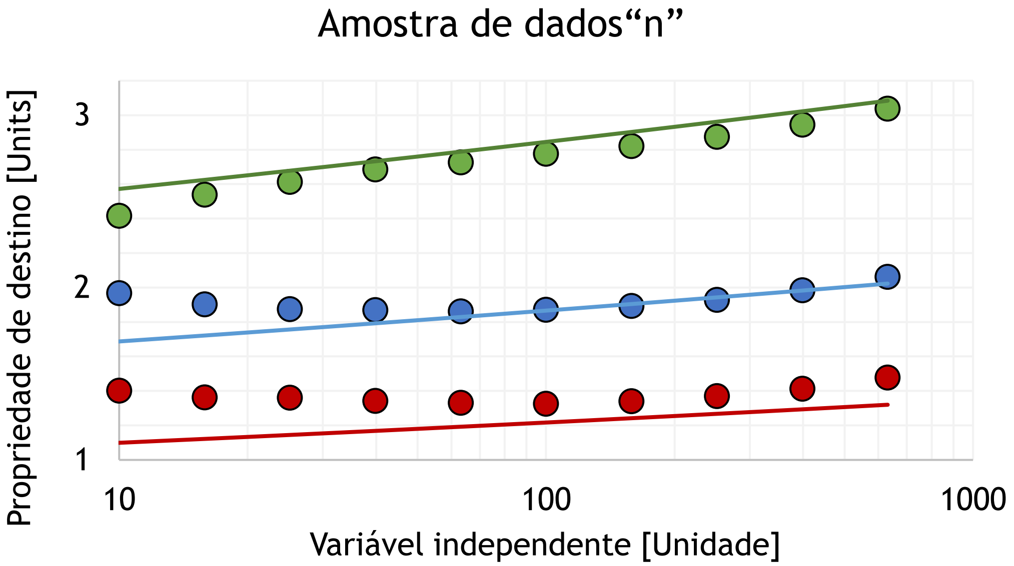
$$D = |Valor\ calculado - Valor\ real|[Unidades]$$

   II. Calcular o desvio médio:

$$DP = \frac{\sum D}{Número\ de\ datos}[Unidades]$$

Critérios de avaliação de regressão:

5. Se necessário, avalie uma amostra dos dados para comparar os gráficos dos dados reais versus os dados estimados com base em qualquer uma das variáveis independentes.



As linhas sólidas são as estimativas.

Os pontos são os dados reais.

Curva	Erro relativo médio %	Desvio médio [Unidade]
Amarelo	2.92	0.079
Azul	3.92	0.075
Verde	11.99	0.165
Média =	6.28	0.11

Exemplo de análise dos resultados de Consumo de Gás Natural

1. Reorganizar a base de dados de consumo de gás natural.
2. Coloque os coeficientes de regressão.
3. Coloque o R^2.
4. Aplique o modelo de regressão aos dados.
5. Calcule o erro relativo médio.
6. Calcule o desvio médio.
7. Plotar resultados e valores reais versus a variável independente.

Federal University of Sergipe
SPE Student Chapter

Misael Edgar Zepeda Díaz