



Universidad Nacional de Colombia  
SPE Student Chapter

PETROIntelligence.com

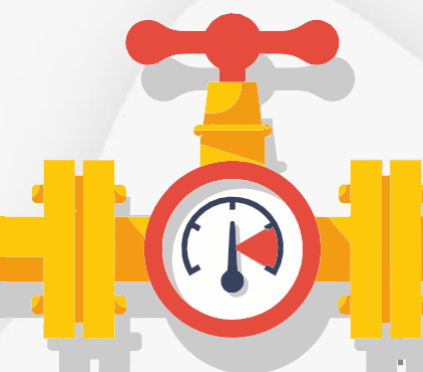
## REGRESIONES APLICADAS A LA INGENIERÍA DE PETRÓLEOS (CON PYTHON)

Misael Edgar Zepeda Díaz

[mzepeda@petrointelligence.com](mailto:mzepeda@petrointelligence.com)

(+52 55 1350 7432)

Agosto 2023





Universidad Nacional de Colombia  
SPE Student Chapter

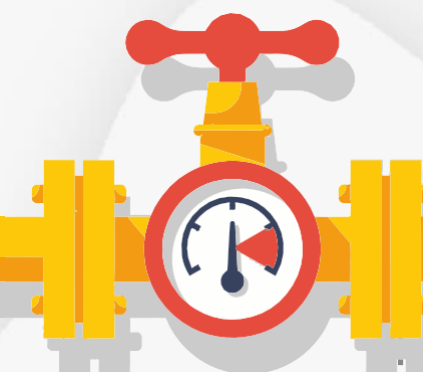
PETROIntelligence.com

## SESIÓN 1 REGRESIONES APLICADAS A LA INGENIERÍA DE PETRÓLEOS (CON PYTHON)

Misael Edgar Zepeda Díaz

[mzepeda@petrointelligence.com](mailto:mzepeda@petrointelligence.com)  
(+52 55 1350 7432)

Agosto 2023

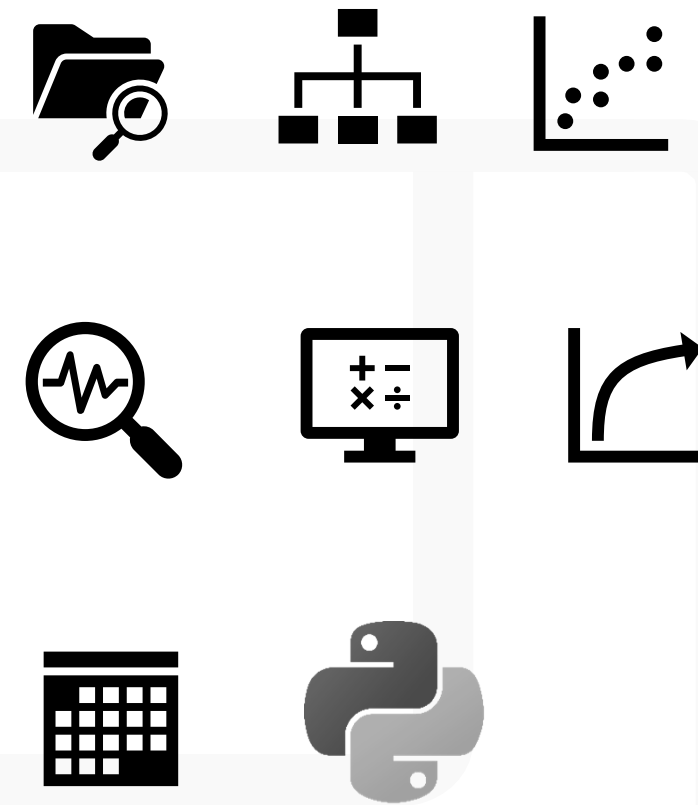




Título	Horas
1. Introducción -----	1
2. Manejo de datos -----	1
3. Python -----	2
4. Regresiones -----	1
5. Análisis de los resultados -----	1
Total =	6

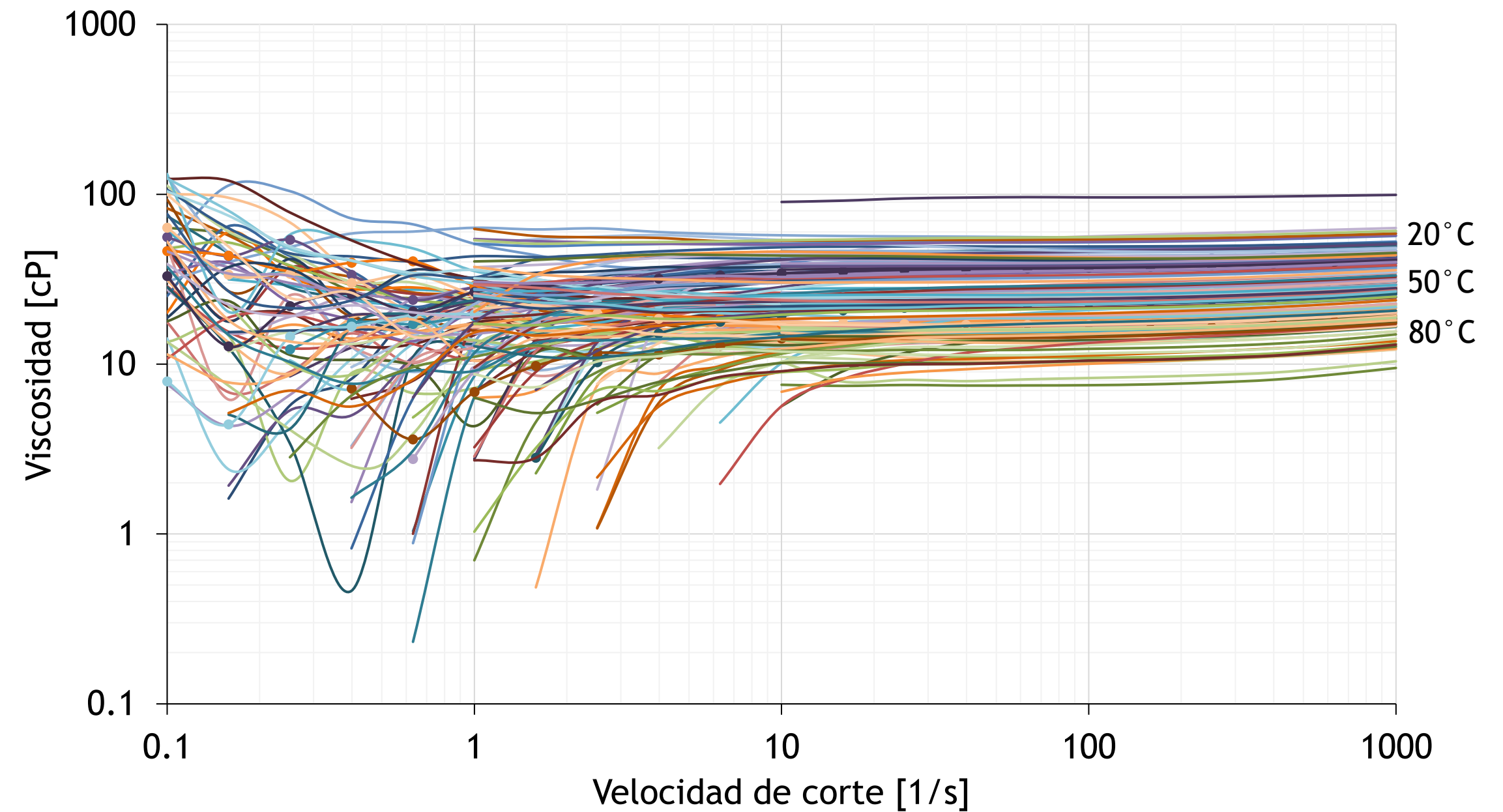
### Enfoque práctico:

- Recolección, organización y visualización de datos.
- Análisis de datos.
- Modelado.
- Regresiones lineales y no lineales.
- Uso de Excel y Python.



## 1. Set de datos:

- a. Viscosidad.
- b. Velocidad de corte.
- c. Temperatura variable.
- d. Densidad variable.

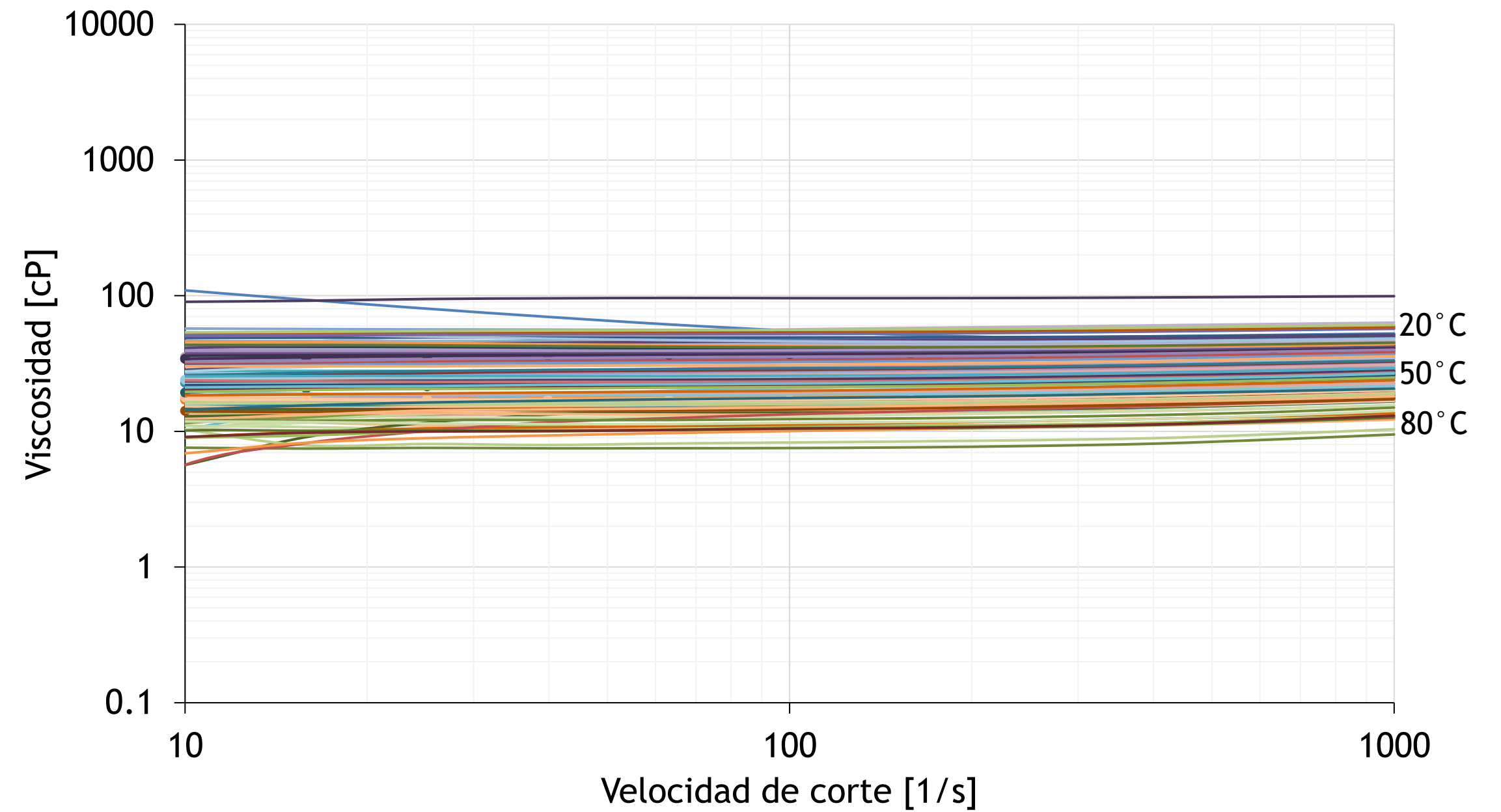


\* Zepeda Díaz, M. 2022. Estudio del comportamiento reológico de muestras de condensado. Tesis de Maestría, UNAM, Programa de Maestría y Doctorado en Ingeniería, Ciudad de México.



## 2. Definición del rango de análisis:

- a. Intervalo de interés.
- b. Tipo de comportamiento.

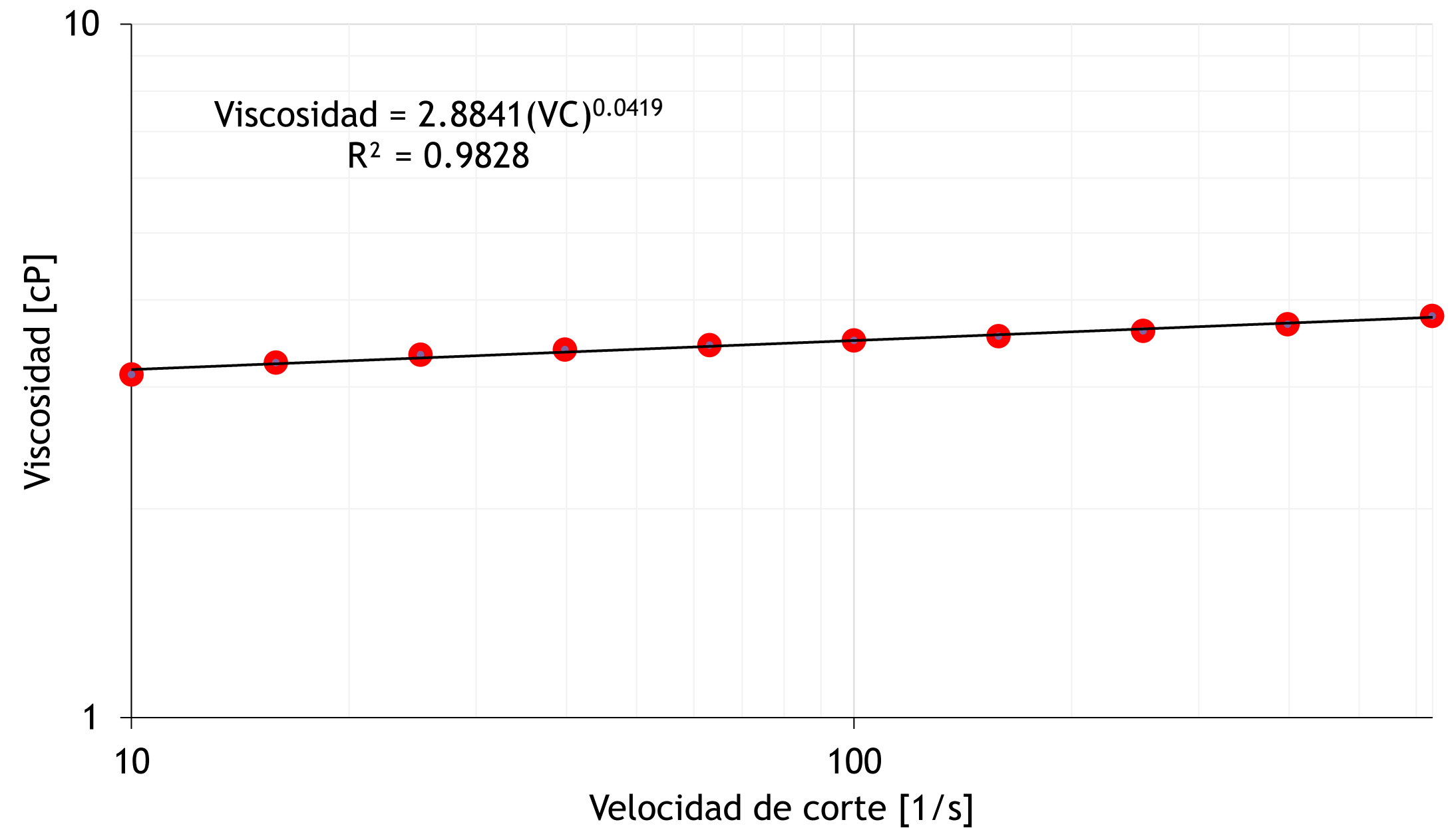


## 3. Definición del modelo:

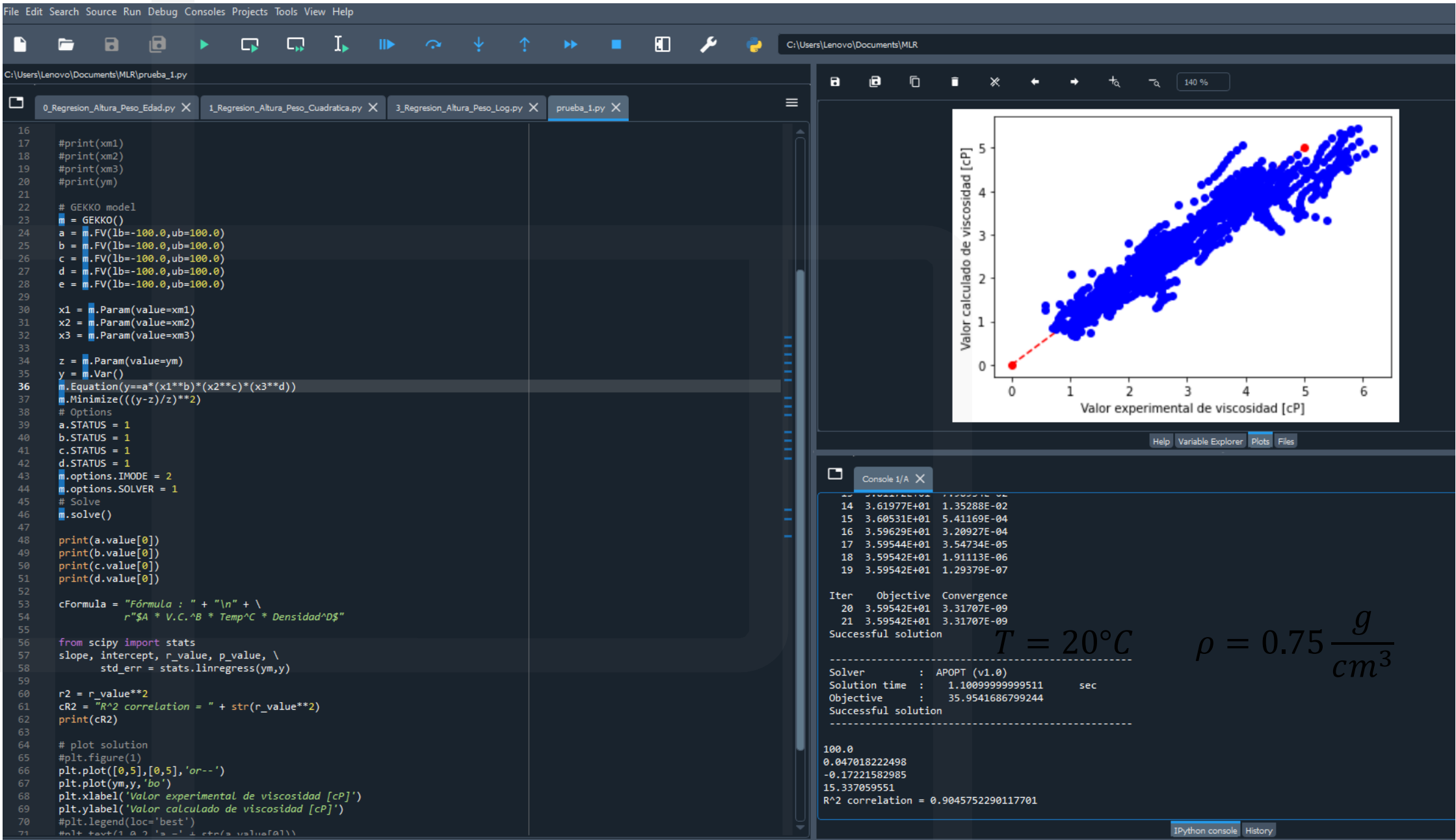
$$\eta = A\dot{\gamma}^B$$

$$T, \rho = cte$$

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



- 4. Regresión con Python:
  - a. Coeficientes de regresión.
  - b. Coeficiente de determinación “ $R^2$ ”.
  - c. Evaluación gráfica de la correlación.





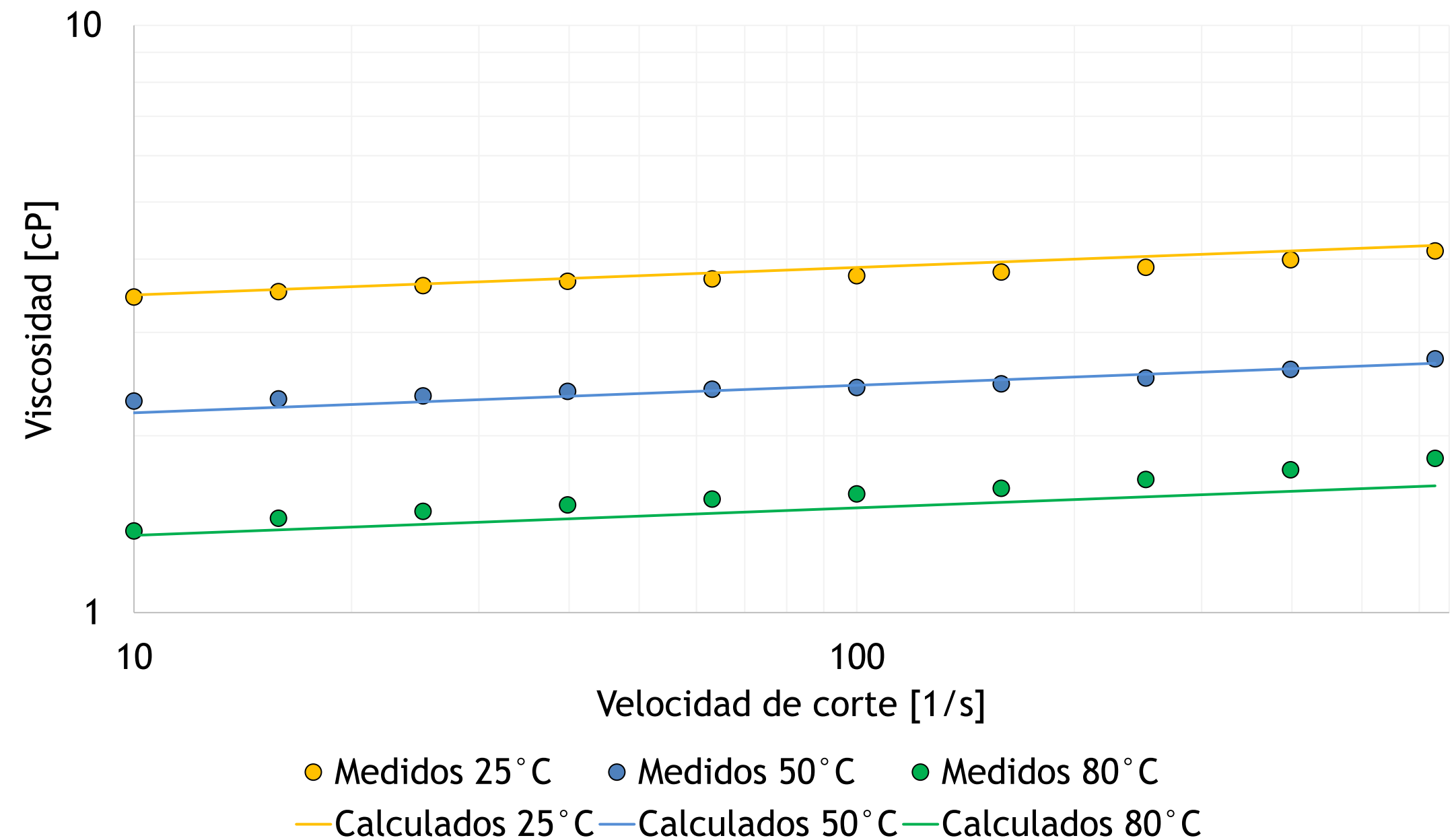
## 5. Evaluación de la regresión:

- Error relativo promedio.
- Desviación promedio.
- Gráfica de datos reales vs calculados

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

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

Muestra n - Ajuste óptimo



## MATERIAL:

<https://classroom.google.com/c/NjE3MzQyMDIzMjI2?cjc=vn44uf6>

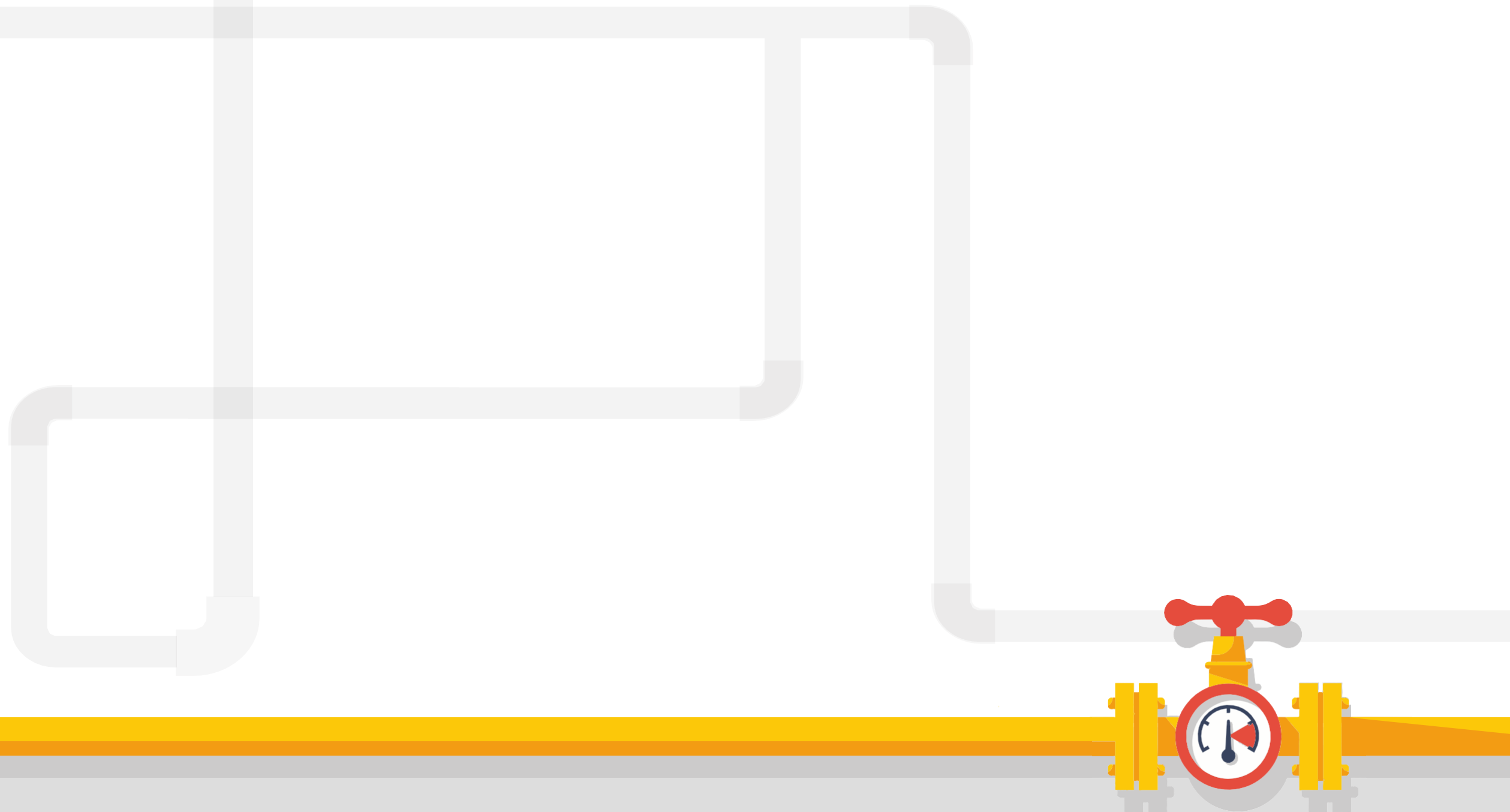
[https://drive.google.com/drive/folders/185NpRh1fPf\\_LW3pGsm7W-mlvmoPcqJjC?usp=drive\\_link](https://drive.google.com/drive/folders/185NpRh1fPf_LW3pGsm7W-mlvmoPcqJjC?usp=drive_link)

# EJERCICIO DE REFORZAMIENTO

---

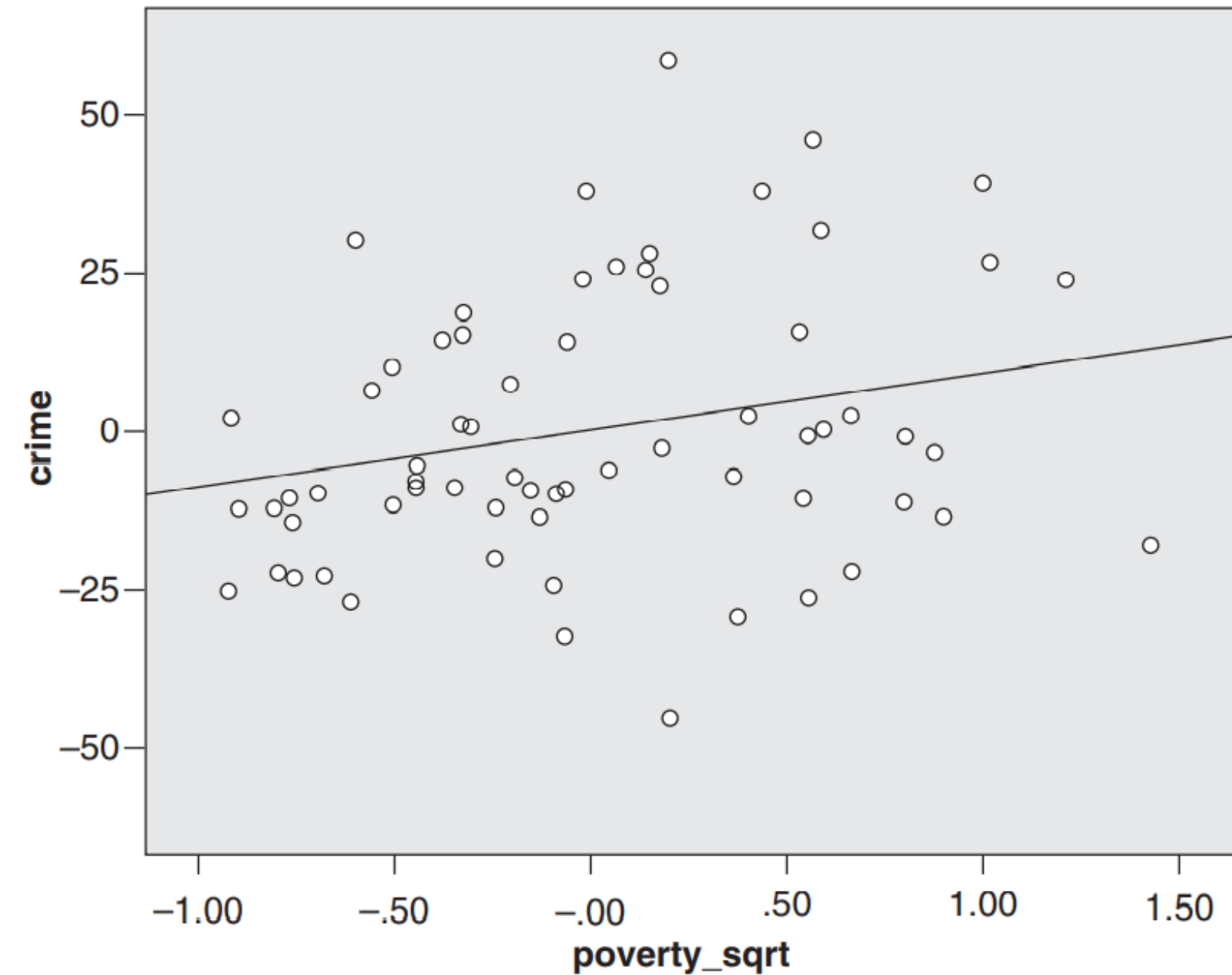
Quiz #1:

<https://acortar.link/i9wtPG>



## Regresión:

Es una herramienta estadística para estimar las relaciones entre una variable dependiente y una o varias variables independientes.

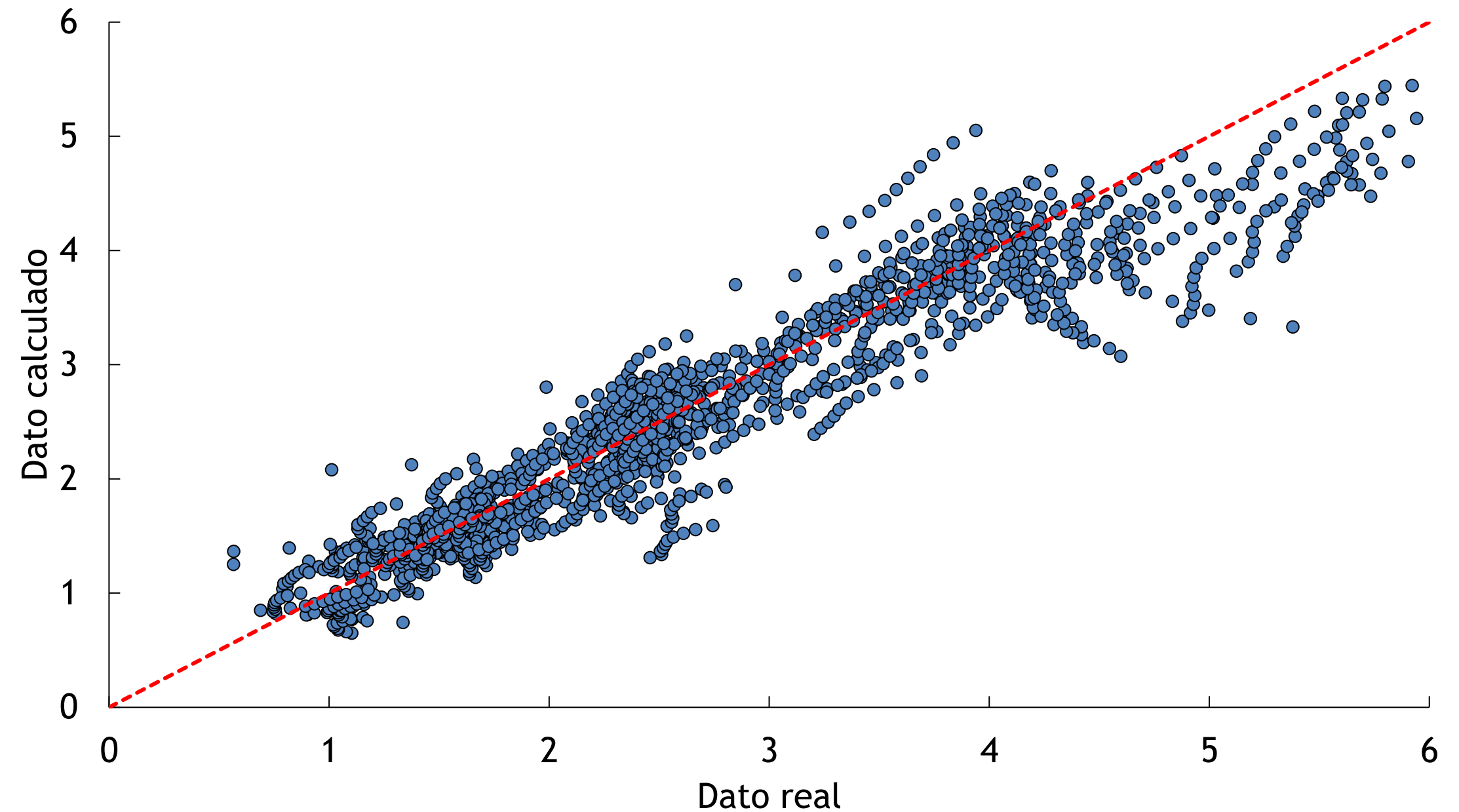


Yang, Qinghua. (2017). Regression. 10.1007/978-3-319-32001-4\_174-1.



## Tipos de regresiones:

1. Regresión lineal.
  - a. Simple.
  - b. Múltiple.
2. Regresión no lineal.
  - a. Simple.
  - b. Múltiple.



## 1. Regresión lineal.

### a. Simple.

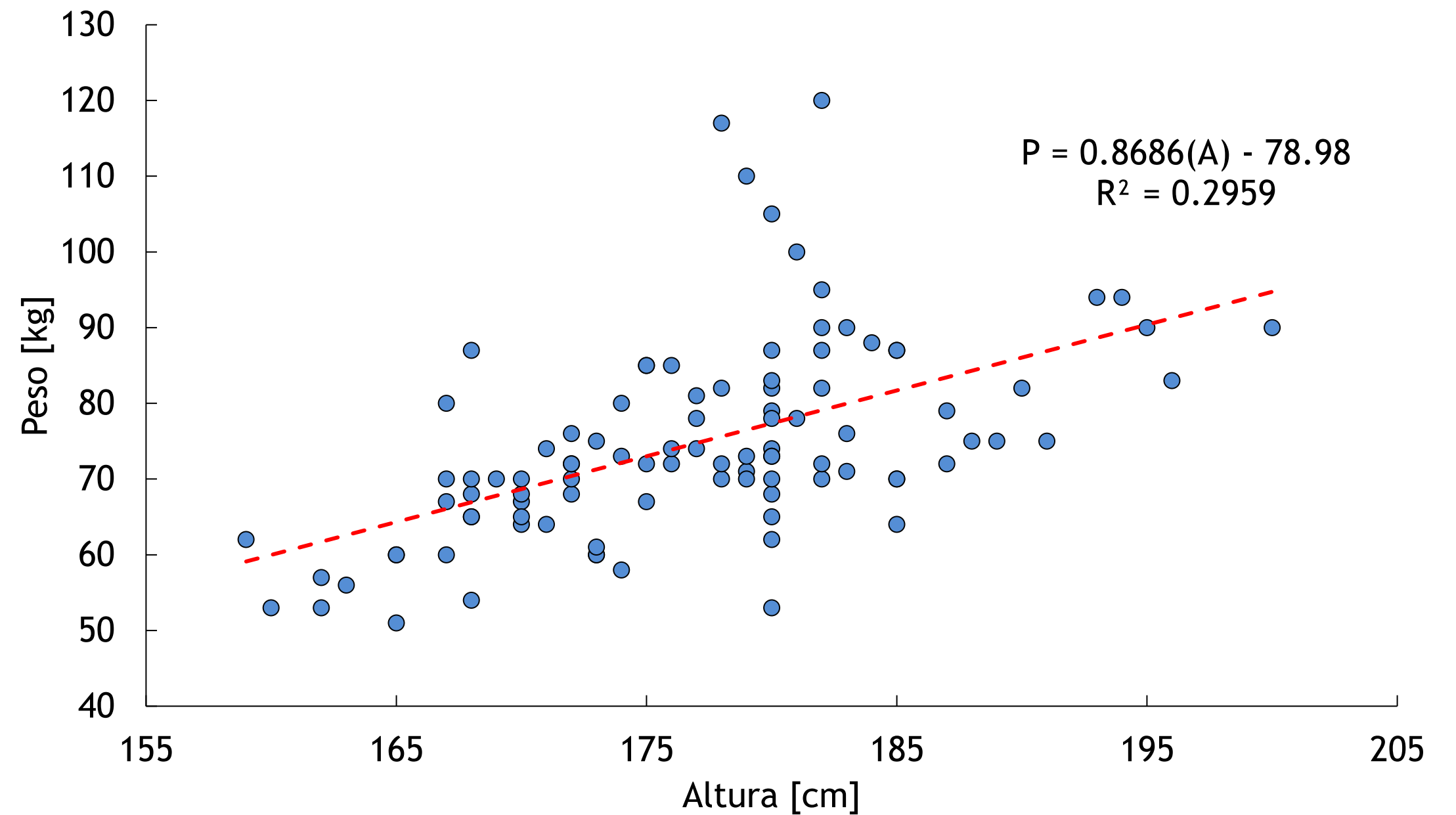
$$y = \alpha + \beta x + \varepsilon$$

$y$ : Variable objetivo.

$x$ : Variable independiente.

$\alpha, \beta$ : Coeficientes de regresión.

$\varepsilon$ : Error.



**#Constantes = #variables independientes + 1**

## 1. Regresión lineal.

### b. Múltiple.

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

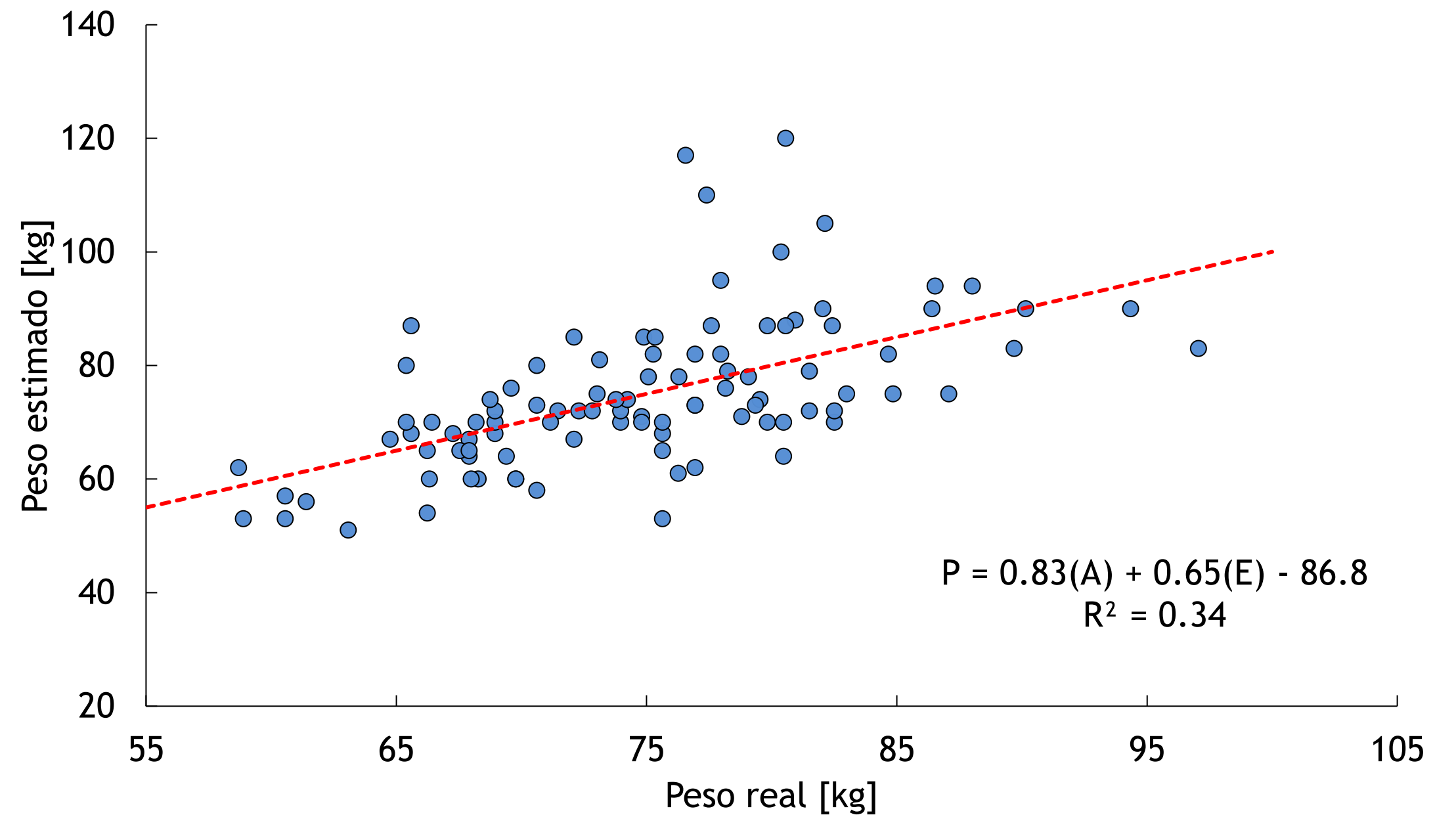
$y$ : Variable objetivo.

$x_1, x_2, \dots, x_k$ : Variable independiente.

$\beta_1, \beta_2, \dots, \beta_k$ : Coeficientes de regresión.

$\varepsilon$ : Error.

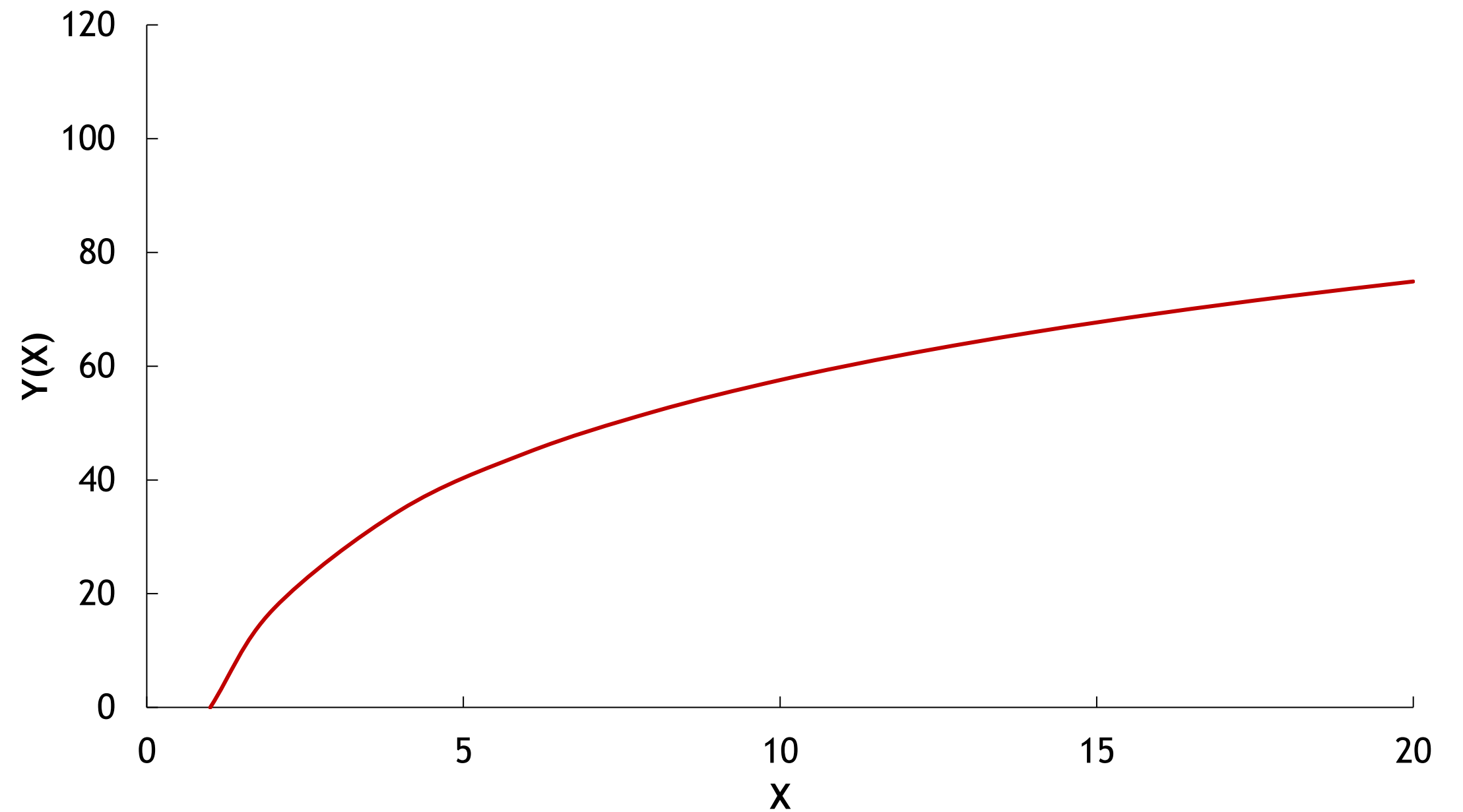
$k$ : K-ésimo término.



## 2. Regresión no lineal.

- Cuadrática.
- Potencial.
- Logarítmica.

¿?

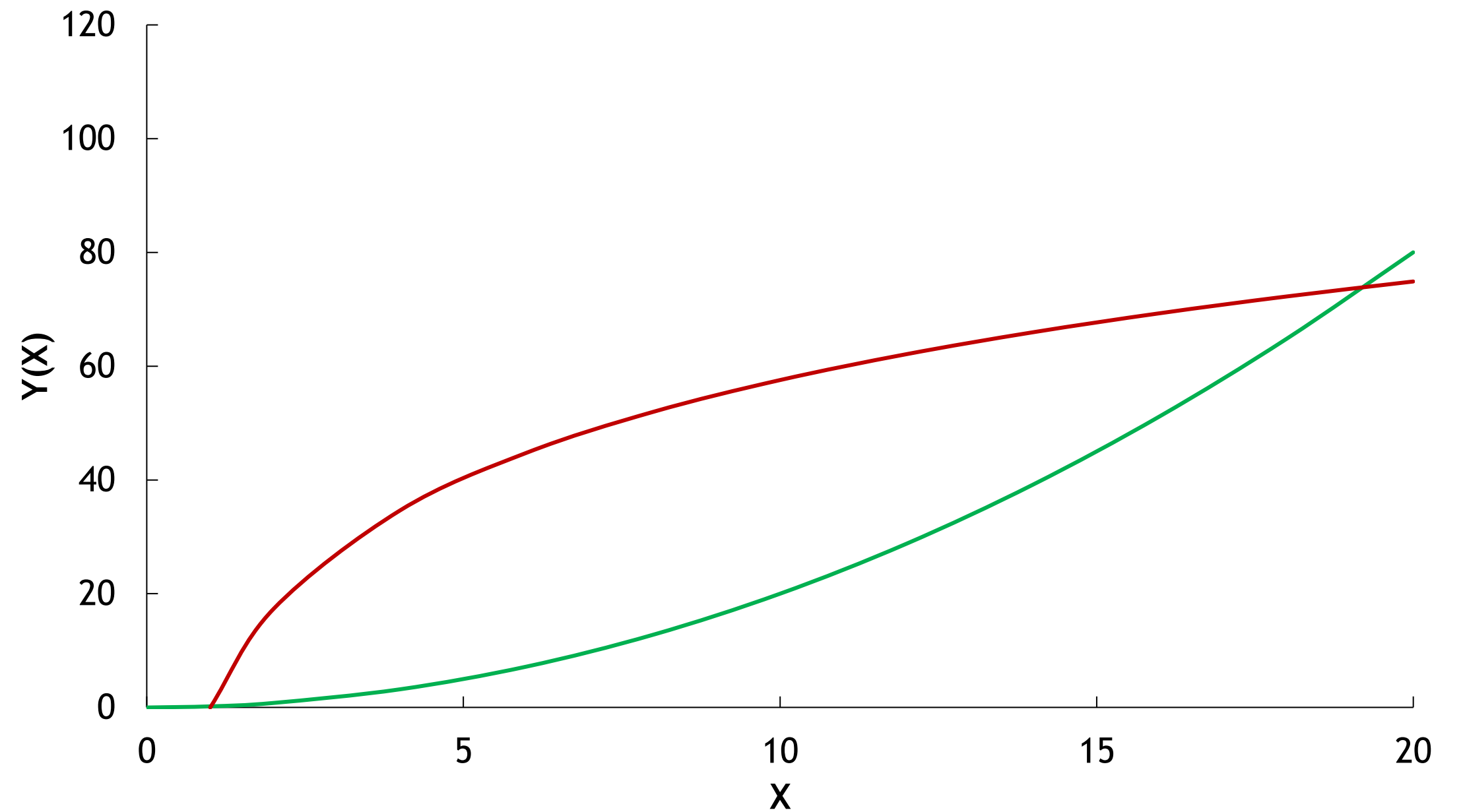




## 2. Regresión no lineal.

- Cuadrática.
- Potencial.
- Logarítmica.

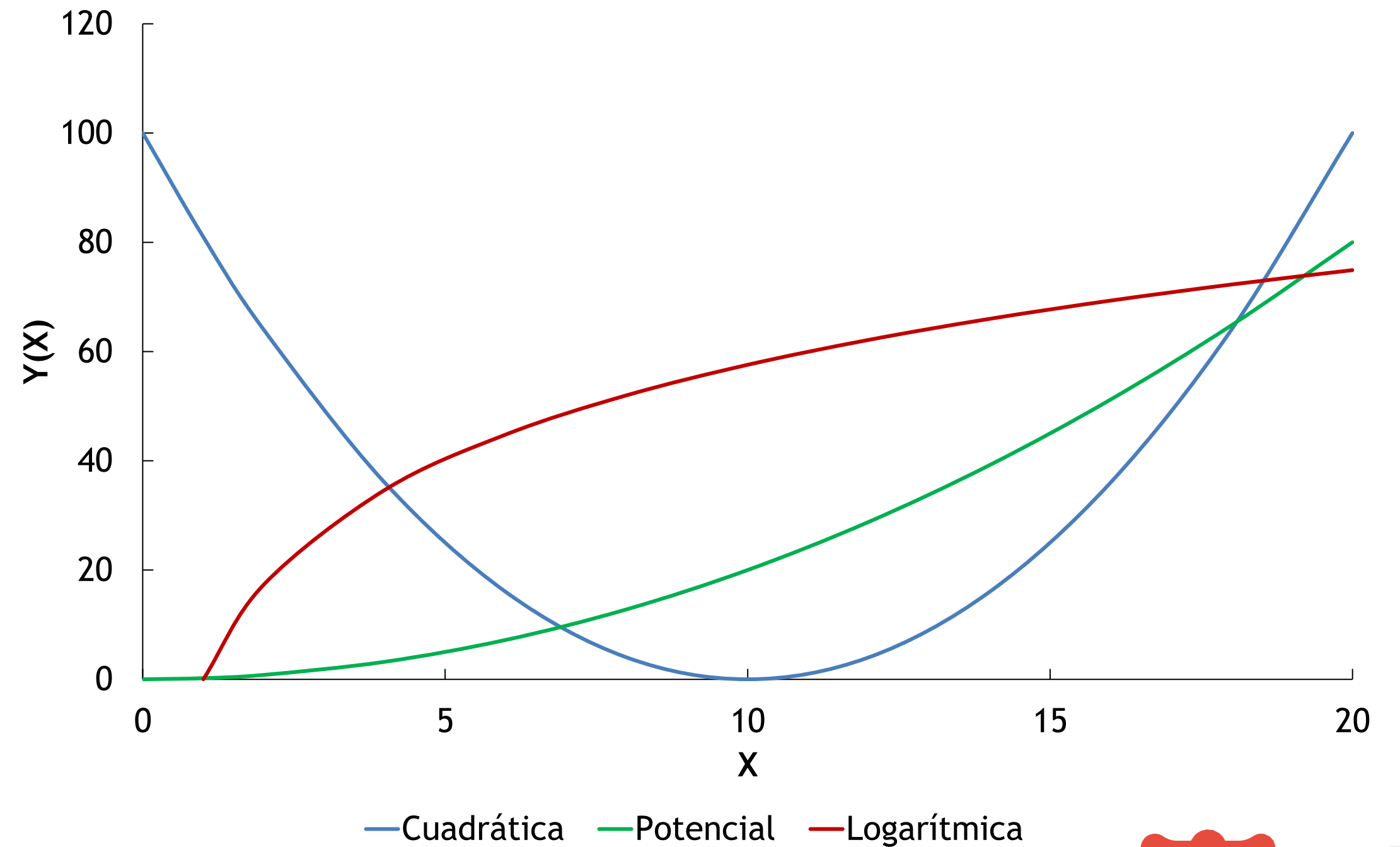
¿?



## 2. Regresión no lineal.

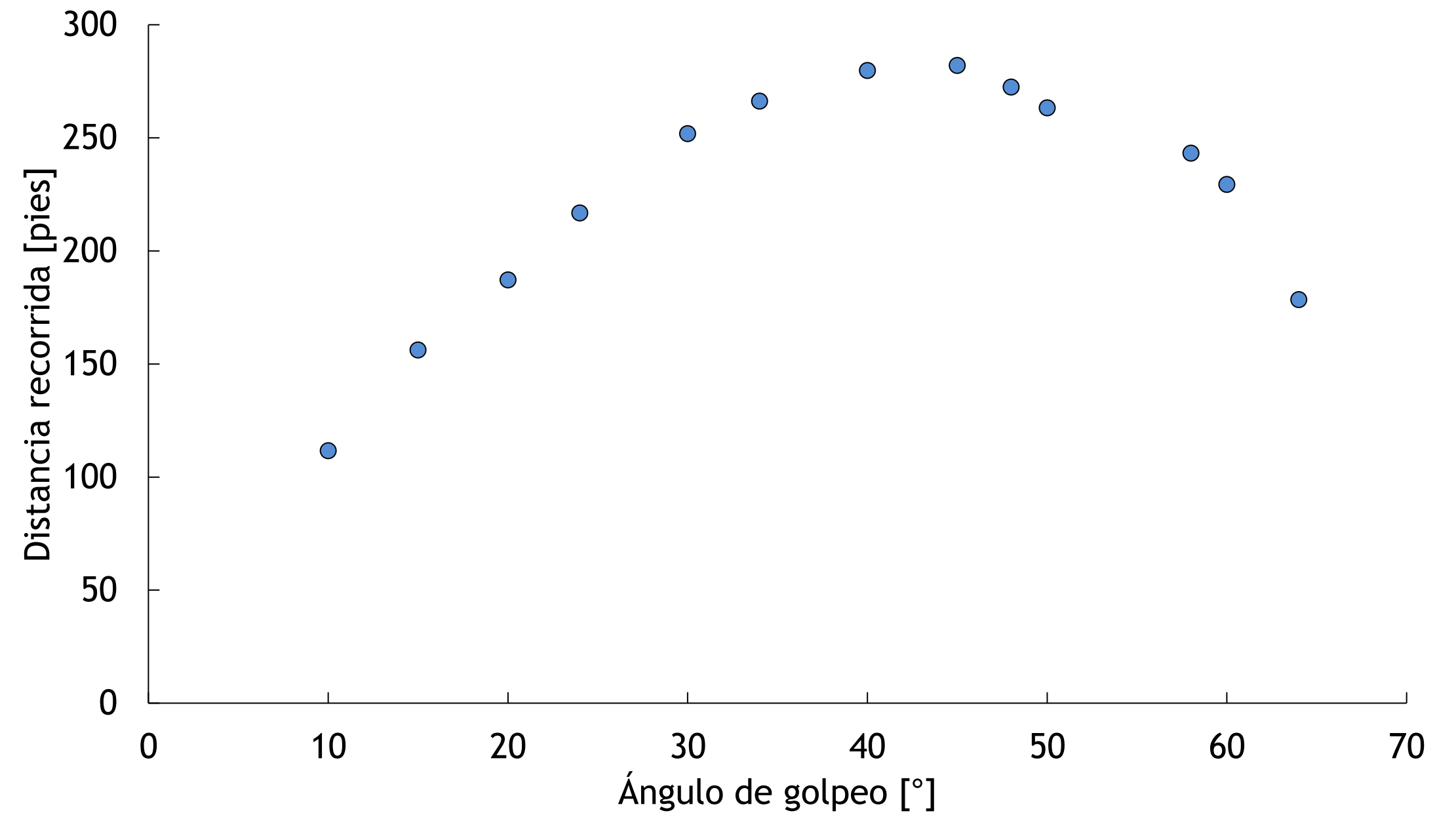
- Cuadrática.
- Potencial.
- Logarítmica.

¿?



## 2. Regresión no lineal.

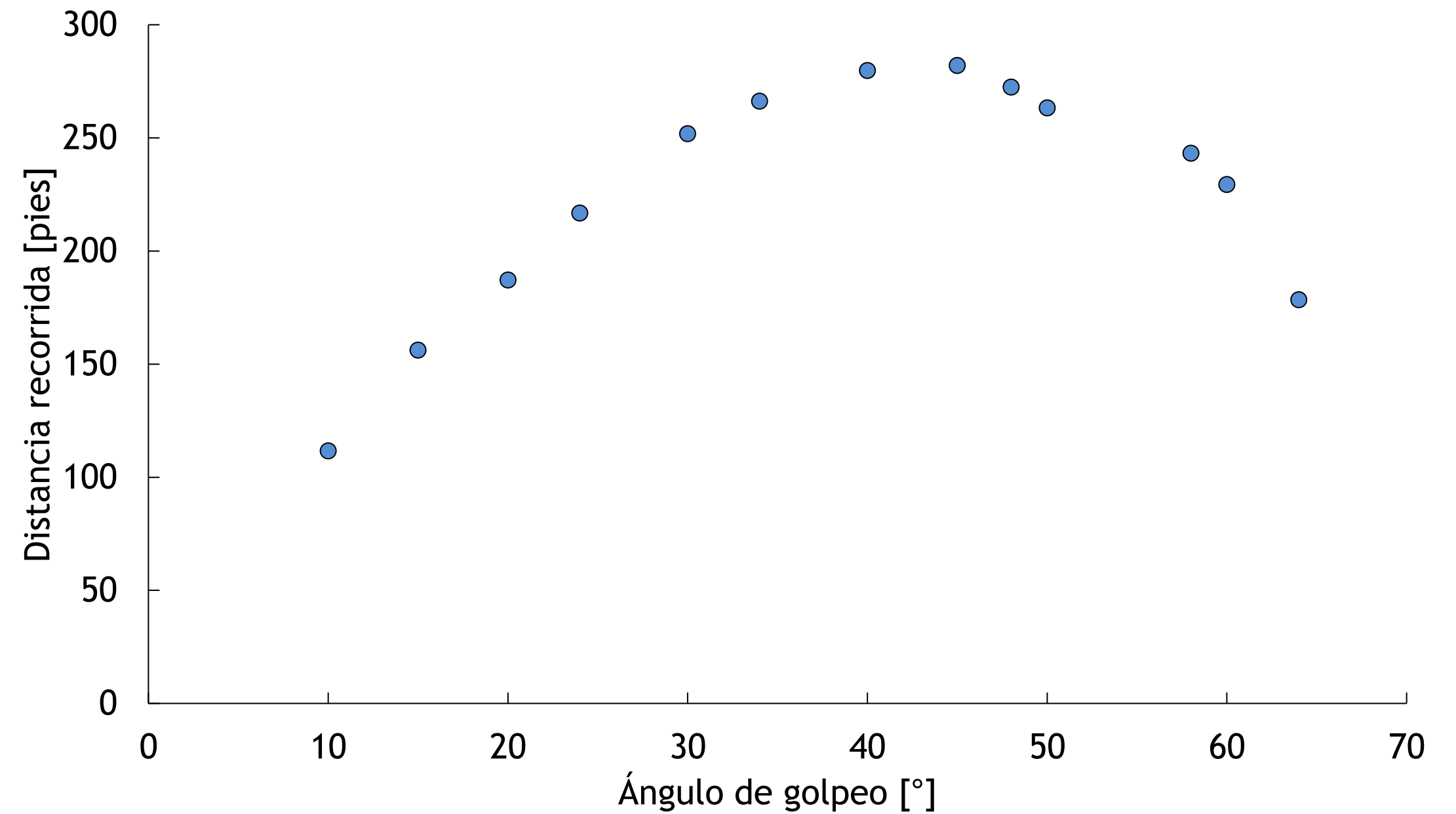
- ¿?



## 2. Regresión no lineal.

- Cuadrática.

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

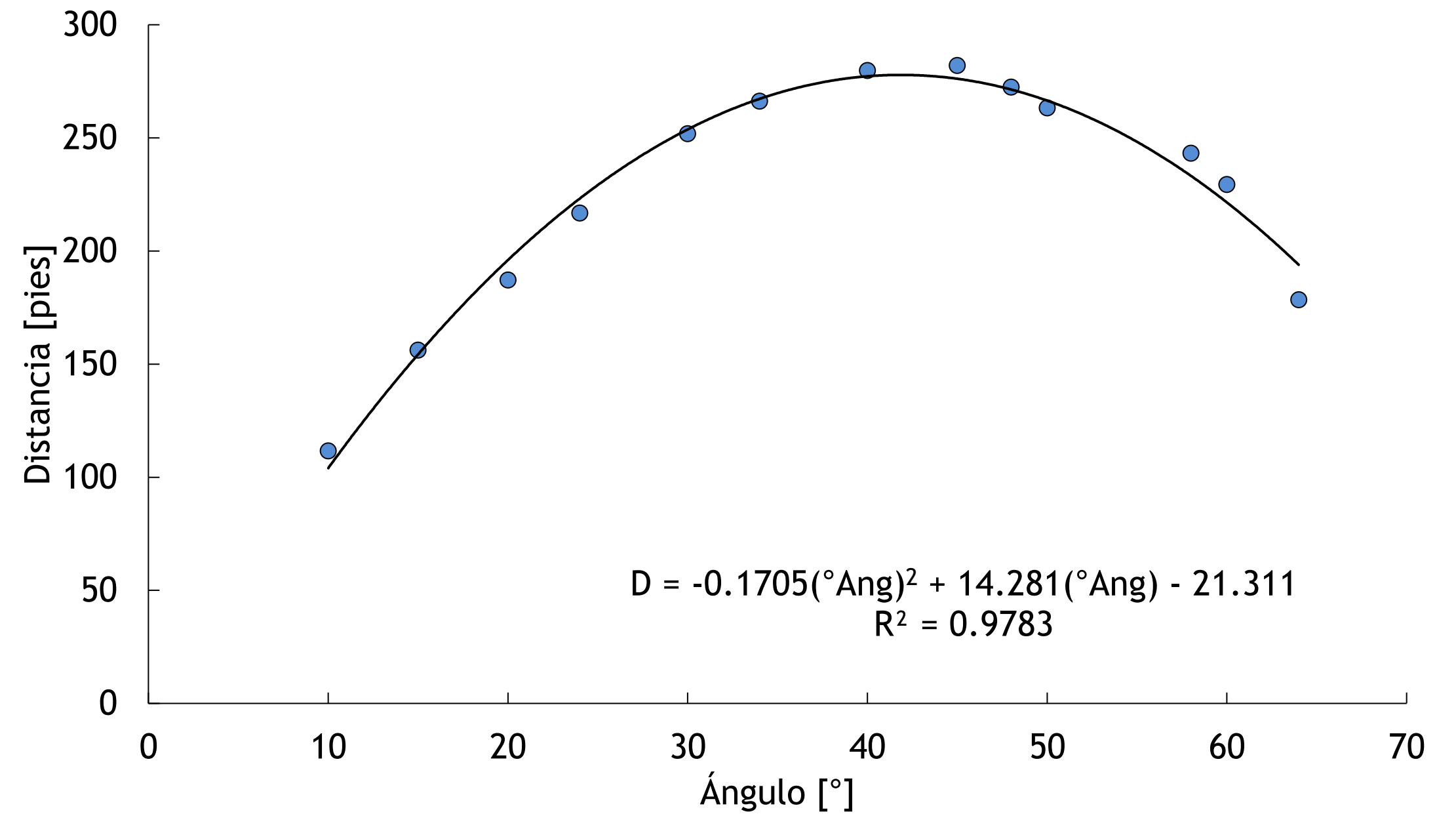




## 2. Regresión no lineal.

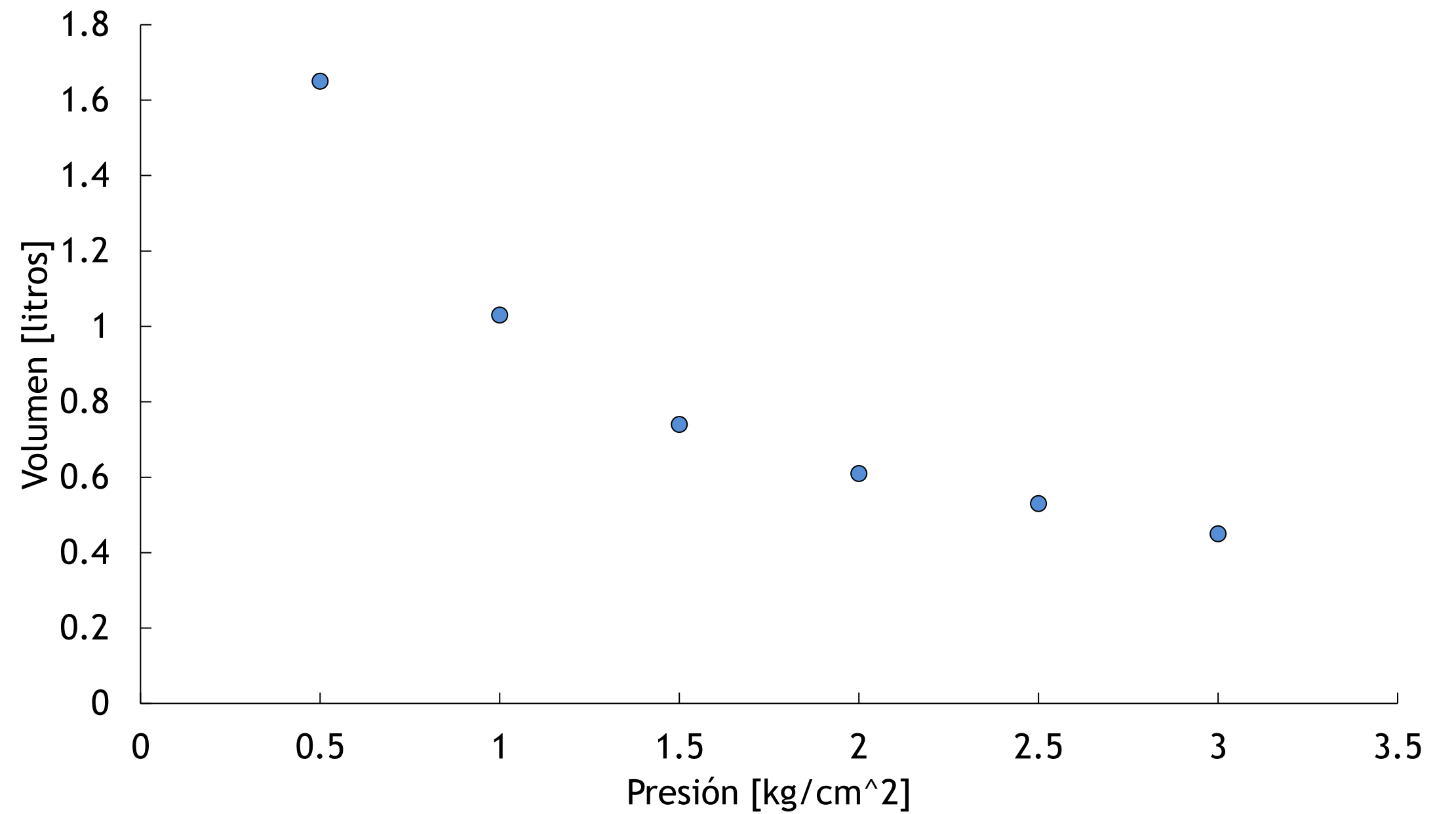
- Cuadrática.

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



## 2. Regresión no lineal.

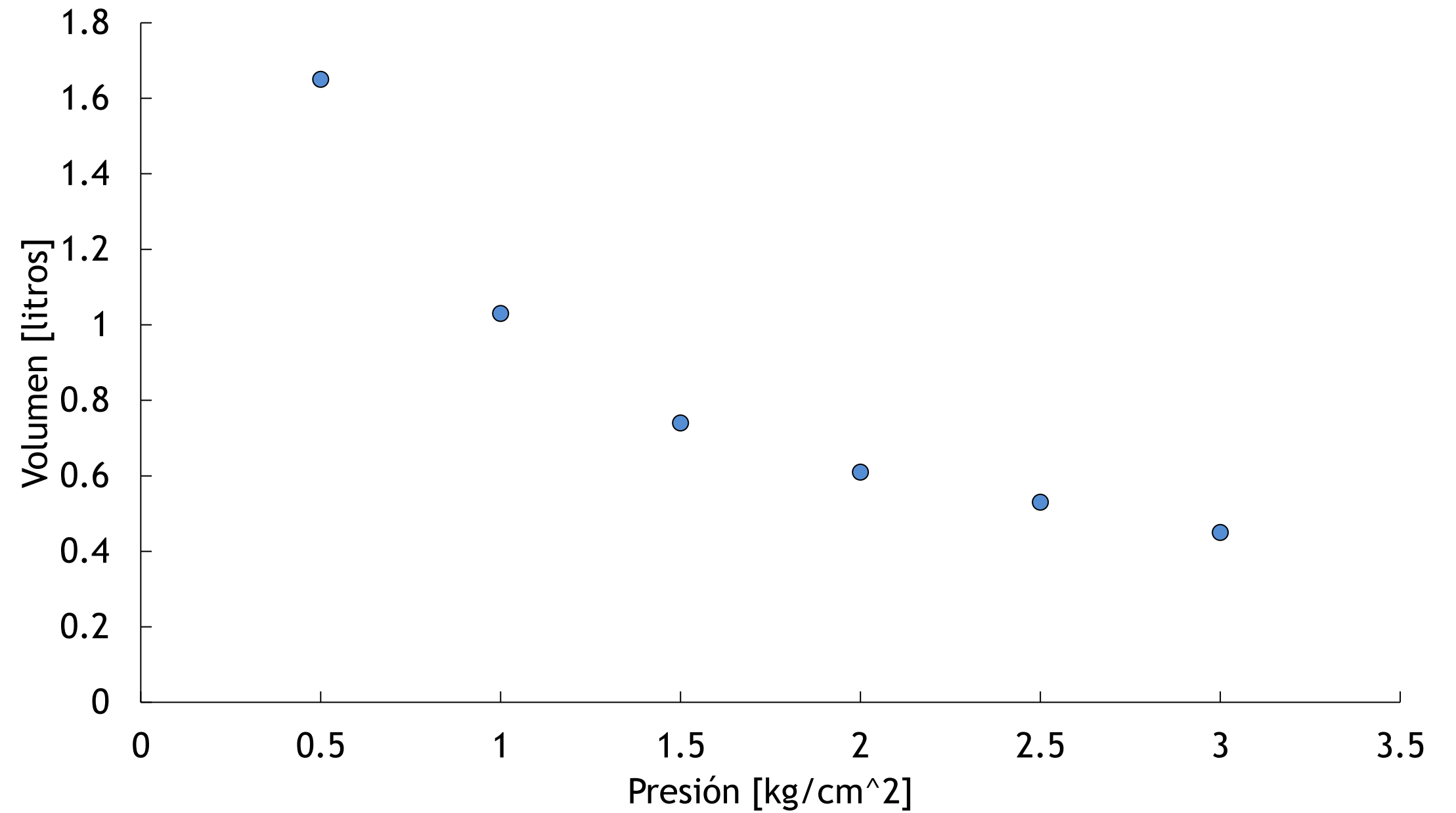
- ¿?



## 2. Regresión no lineal.

- Potencial.

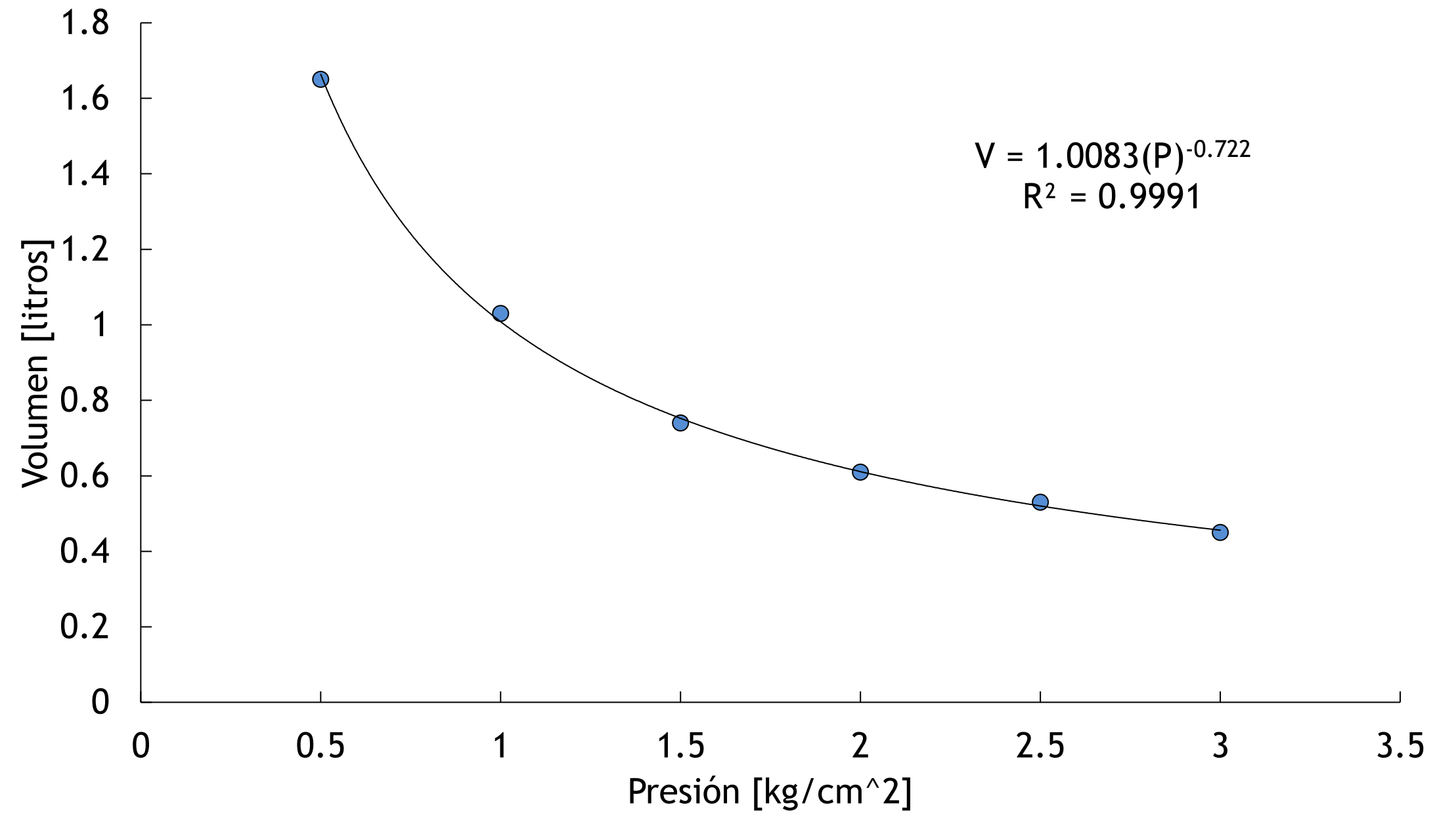
$$y = \alpha(x_1)^\beta + \varepsilon$$



## 2. Regresión no lineal.

- Potencial.

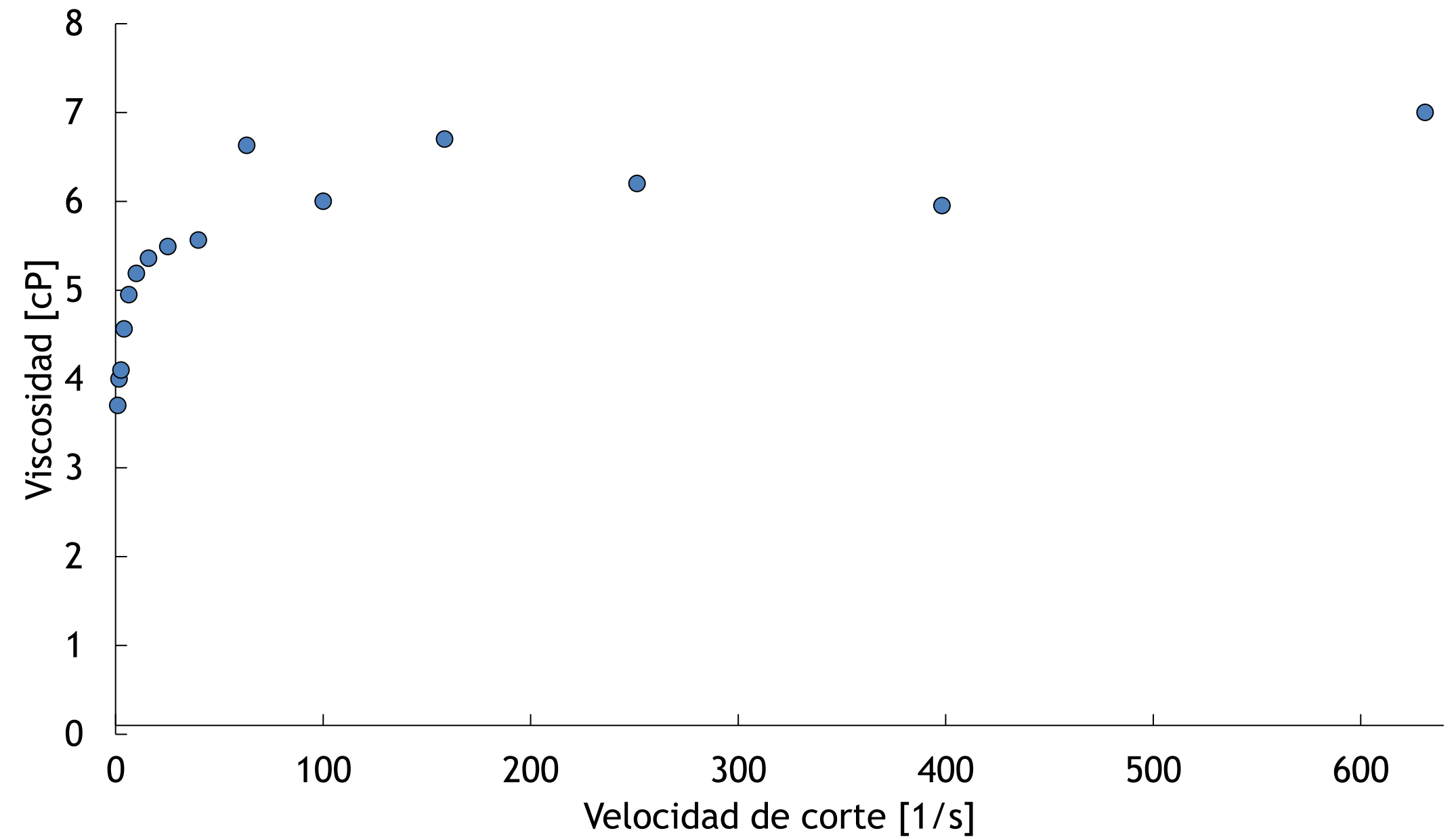
$$y = \alpha(x_1)^\beta + \varepsilon$$





## 2. Regresión no lineal.

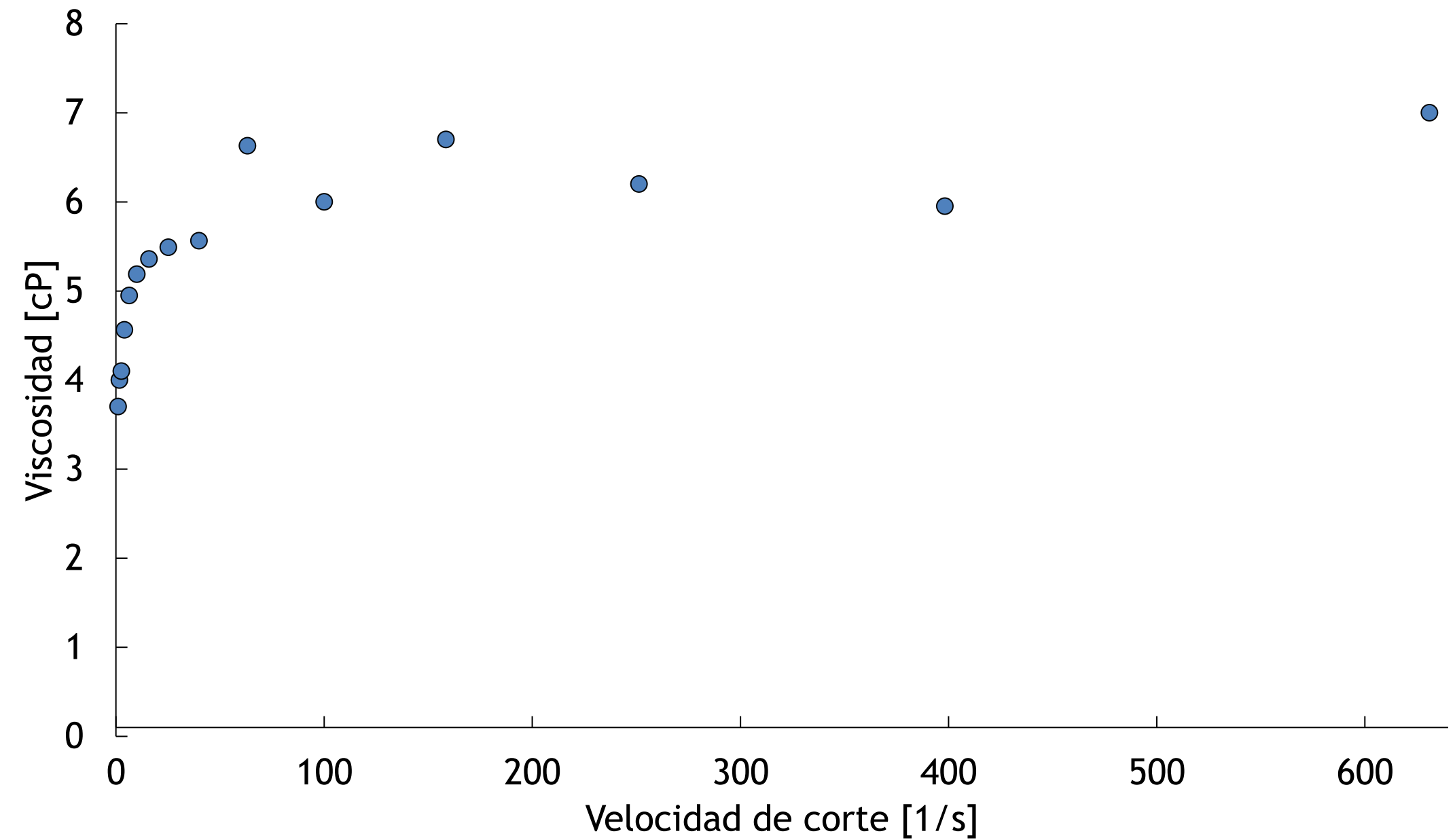
- ¿?



## 2. Regresión no lineal.

- Logarítmica.

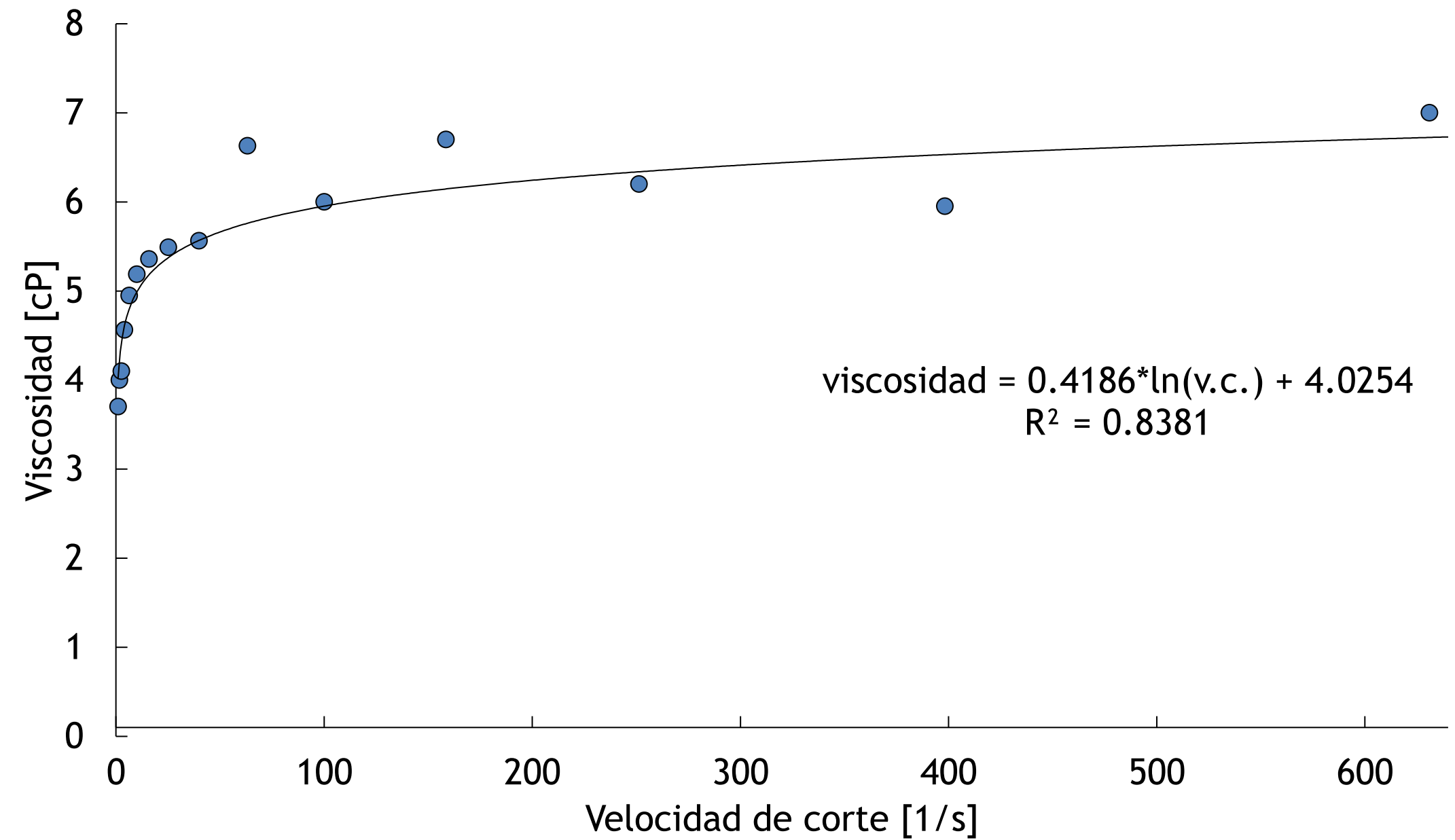
$$y = \alpha \ln(x) + \varepsilon$$



## 2. Regresión no lineal.

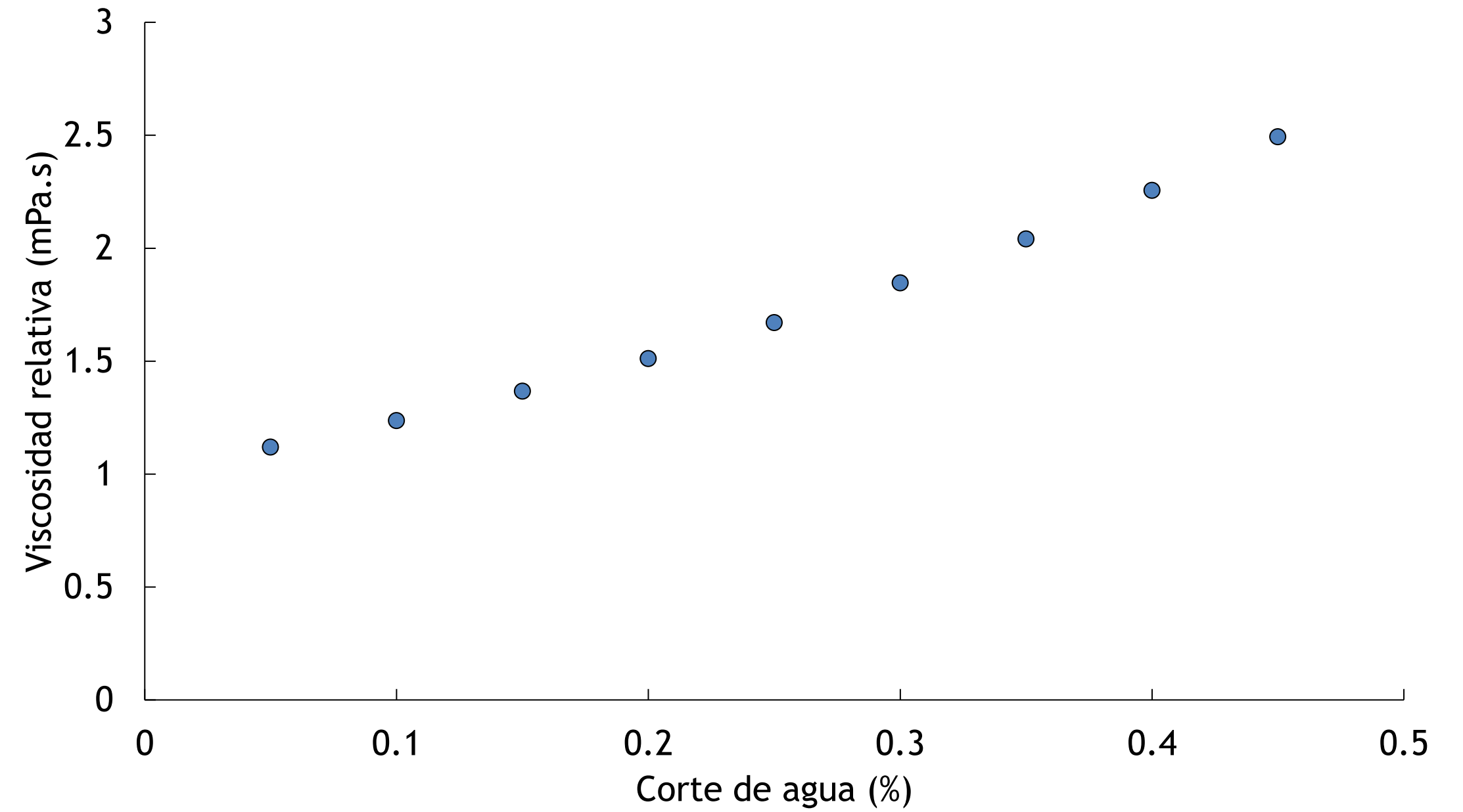
- Logarítmica.

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



## 2. Regresión no lineal.

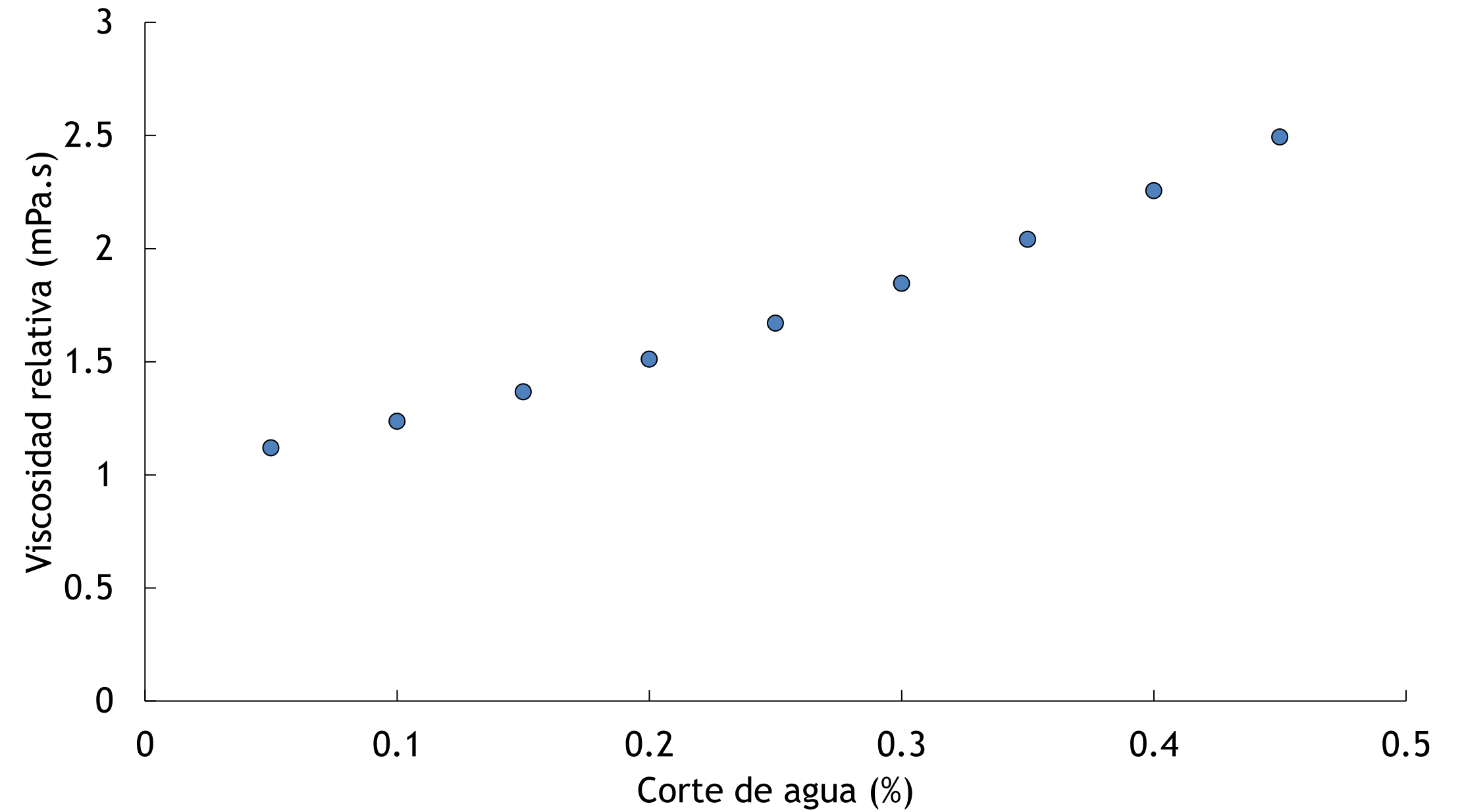
- ¿?



## 2. Regresión no lineal.

- Exponencial

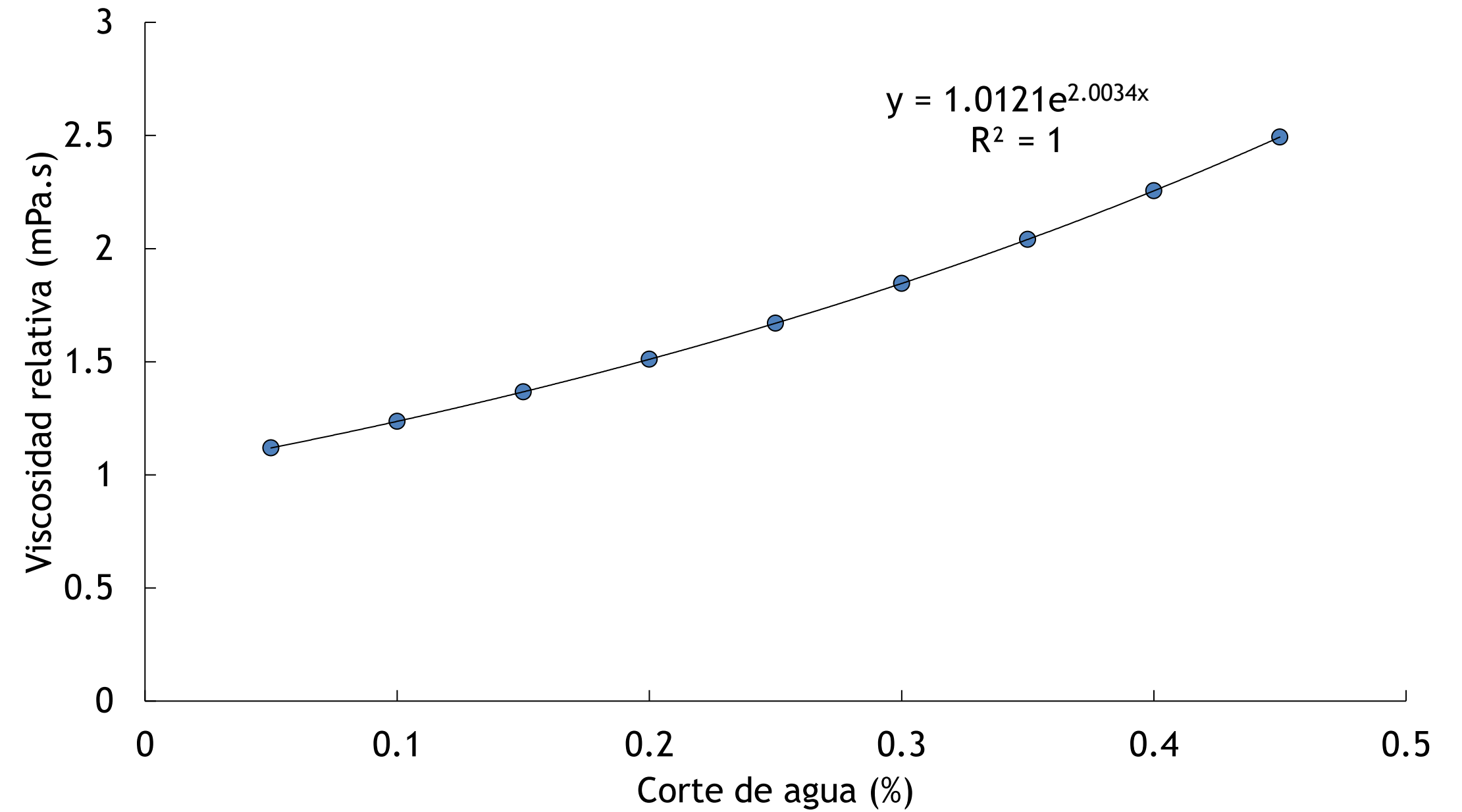
$$\ln(\eta_r) = k_1 + k_2T + k_3\phi + k_4T\phi$$



## 2. Regresión no lineal.

- Exponencial

$$\ln(\eta_r) = k_1 + k_2T + k_3\phi + k_4T\phi$$



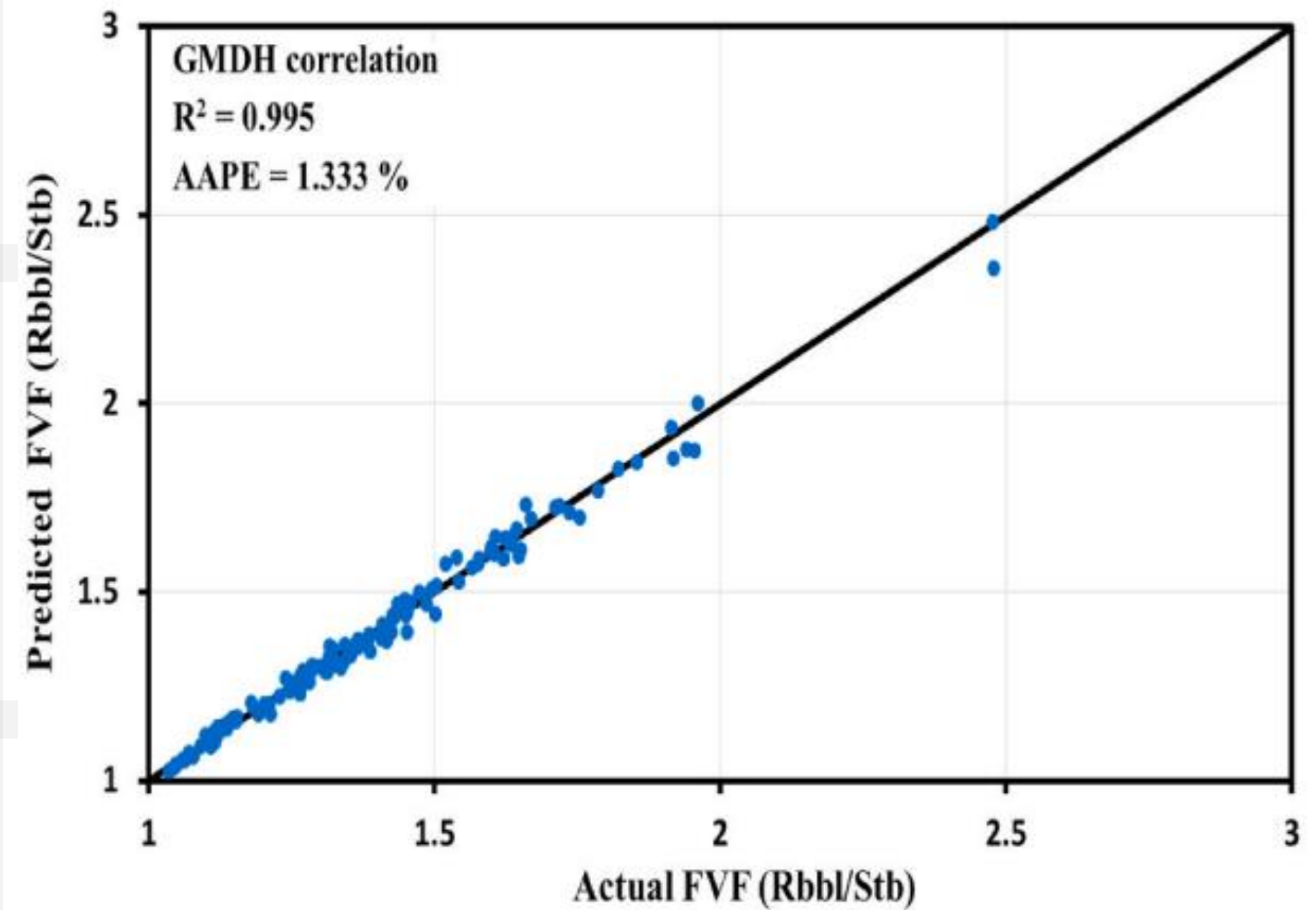


## Predicción del Factor de Volumen del Aceite:

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

$$R_s, T, ^\circ API, \gamma_g$$

\*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*



## Regresión de López-Hernández (2017):

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

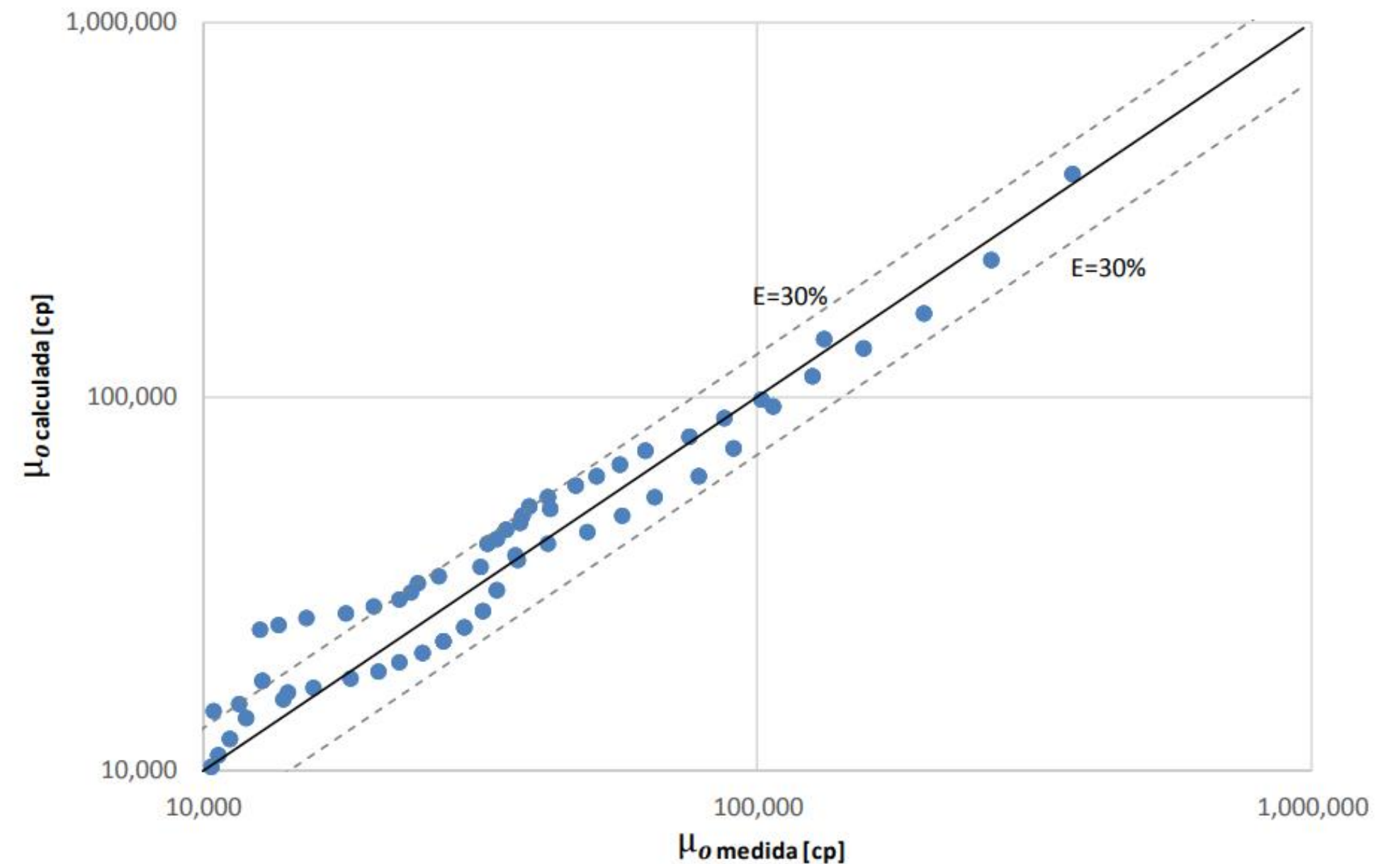
$K_{1,2,3}$ : constantes de regresión

$\mu_o$ : viscosidad del aceite

$T$ : temperatura

$\gamma$ : velocidad de corte

\* López-Hernández O. (2017). Estudio Reológico de Emulsiones de Agua y Aceite Crudo Pesado de Campos Marinos en México. Tesis de maestría. Universidad Nacional Autónoma de México.



## Regresión de López-Pérez (2021):

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

$K_{1,2,3,4,5}$  : constantes de regresión

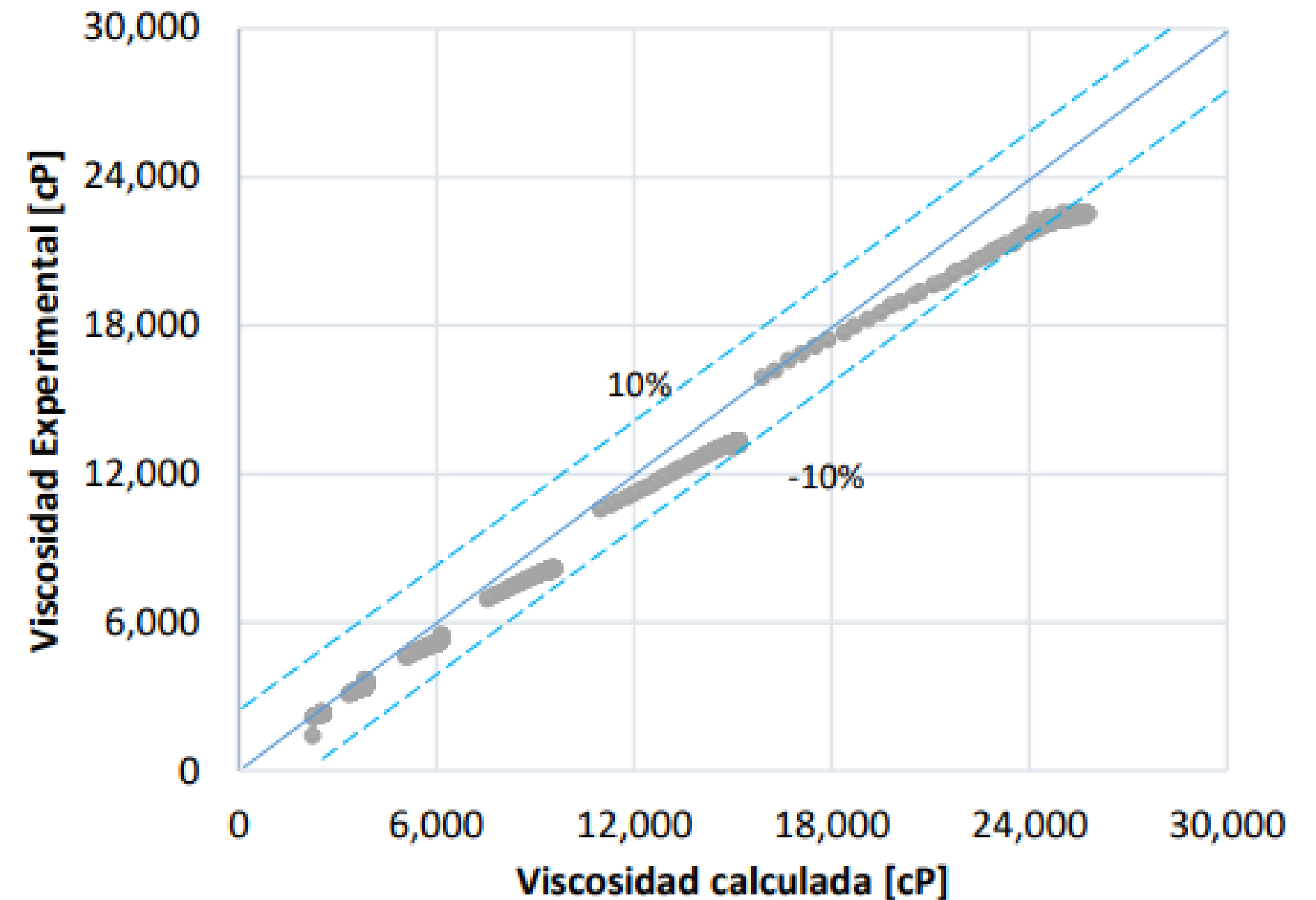
$\mu_r$  : viscosidad del aceite

$\phi$  : corte de agua

$T$  : temperatura

$\gamma$  : velocidad de corte

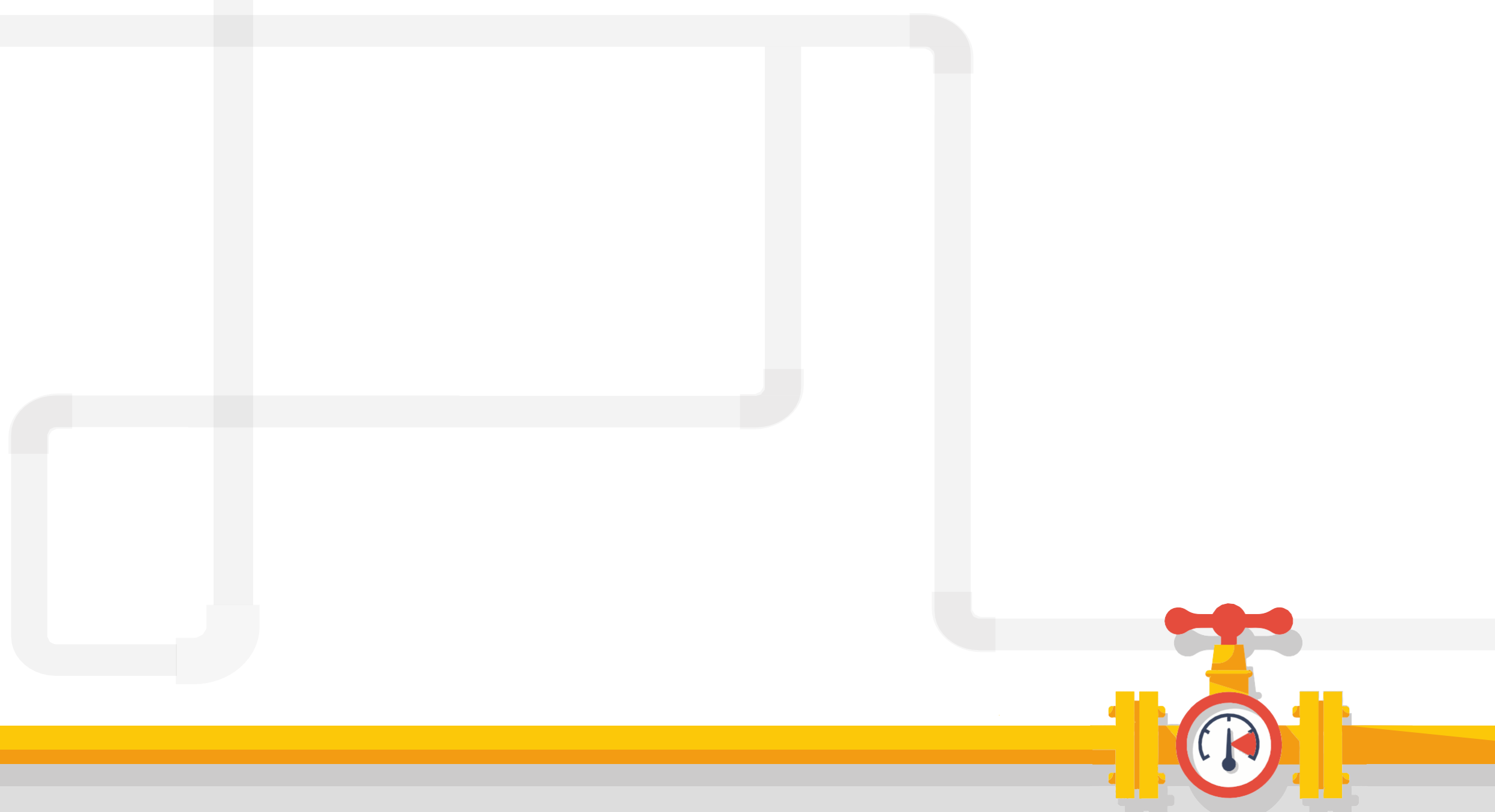
\* López-Pérez V. (2021). Obtención de una correlación para determinar la viscosidad de emulsiones agua-aceite en función de temperatura, contenido de agua y velocidad de corte. Tesis de maestría. Instituto Mexicano del Petróleo.



# EJERCICIO DE REFORZAMIENTO

## Quiz #2:

<https://acortar.link/tganuo>







Universidad Nacional de Colombia  
SPE Student Chapter

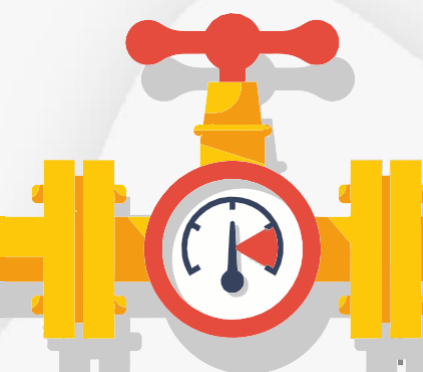
PETROIntelligence.com

## SESIÓN 2 REGRESIONES APLICADAS A LA INGENIERÍA DE PETRÓLEOS (CON PYTHON)

Misael Edgar Zepeda Díaz

[mzepeda@petrointelligence.com](mailto:mzepeda@petrointelligence.com)  
(+52 55 1350 7432)

Agosto 2023

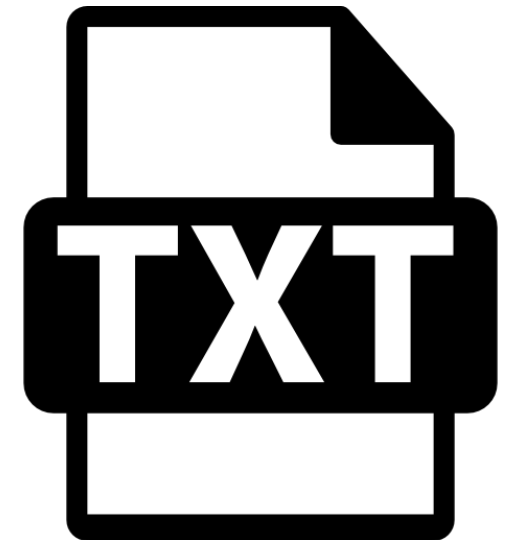
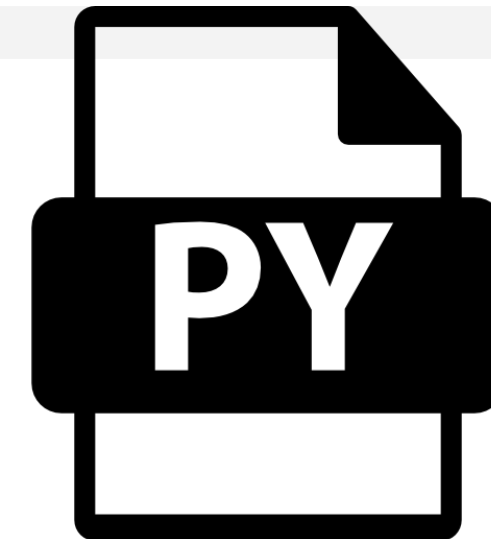


Título	Horas
1. Introducción -----	1
2. Manejo de datos -----	1
3. Python -----	2
4. Regresiones -----	1
5. Análisis de los resultados -----	1
Total =	6



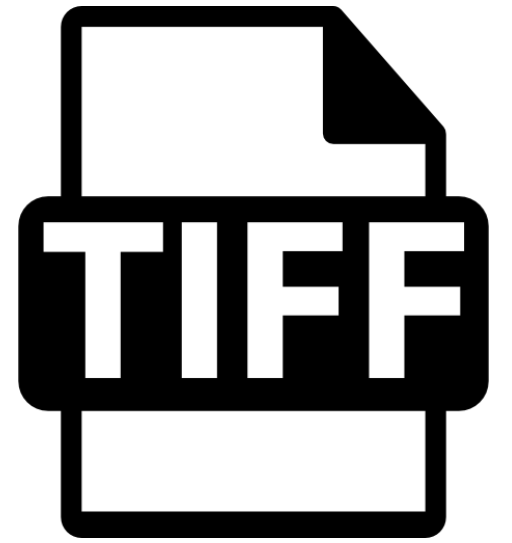
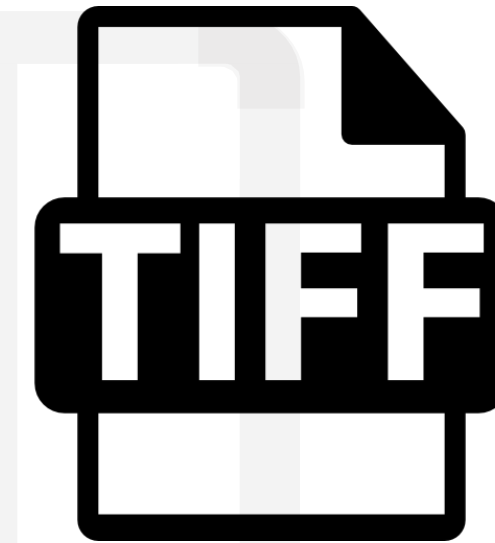
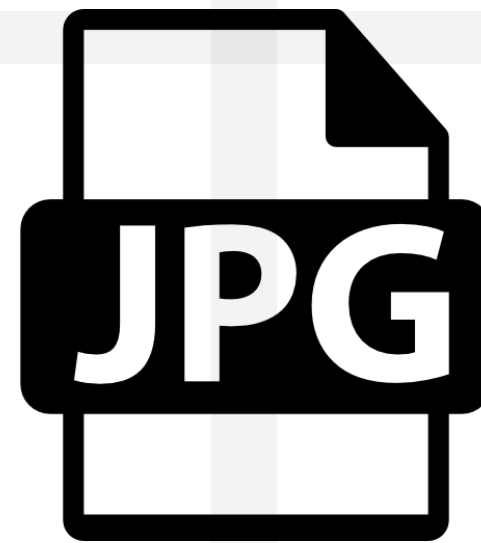
## Formatos que se utilizarán:

1. CSV
2. PY
3. XLSX
4. TXT
5. PDF
6. DOC



### Formatos adicionales que conozcan:

1. JPG
2. LAS (datos de registros eléctricos)
3. RAR
4. TIFF (georreferencia)
5. PNG
6. HTML
7. M
8. CPP



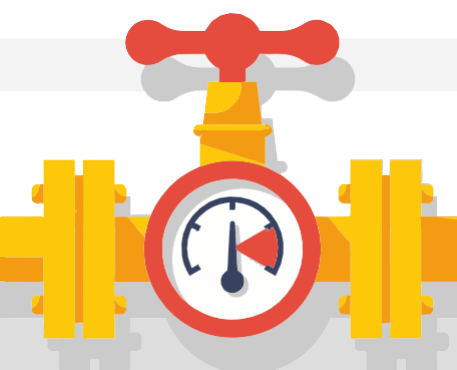
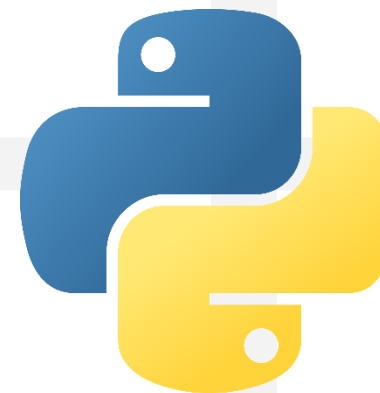
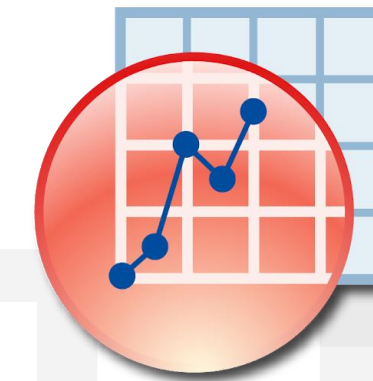
## Visualización:

### 1. Excel:

- I. Tradicional.
- II. Tablas dinámicas.

### 2. Otras herramientas:

- I. OriginLab.
- II. Python (Matplotlib, Seaborn).
- III. Power BI.
- IV. Tableau.
- V. Google Charts.



## Fuentes de datos:

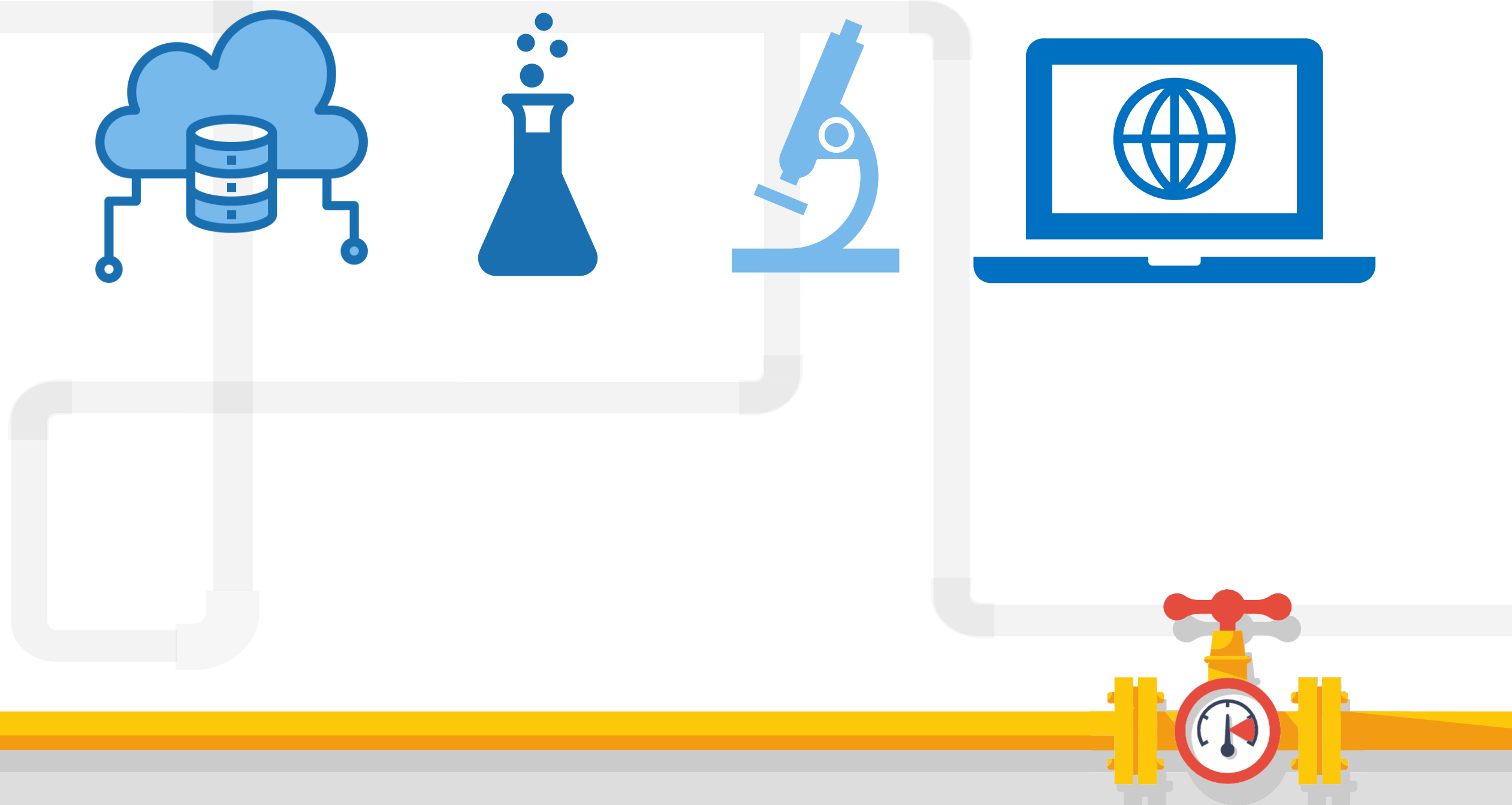
- a. Externas.
- b. Internas.



## Fuentes de datos:

### a. Externas:

- <https://regressit.com>



### Fuentes de datos:

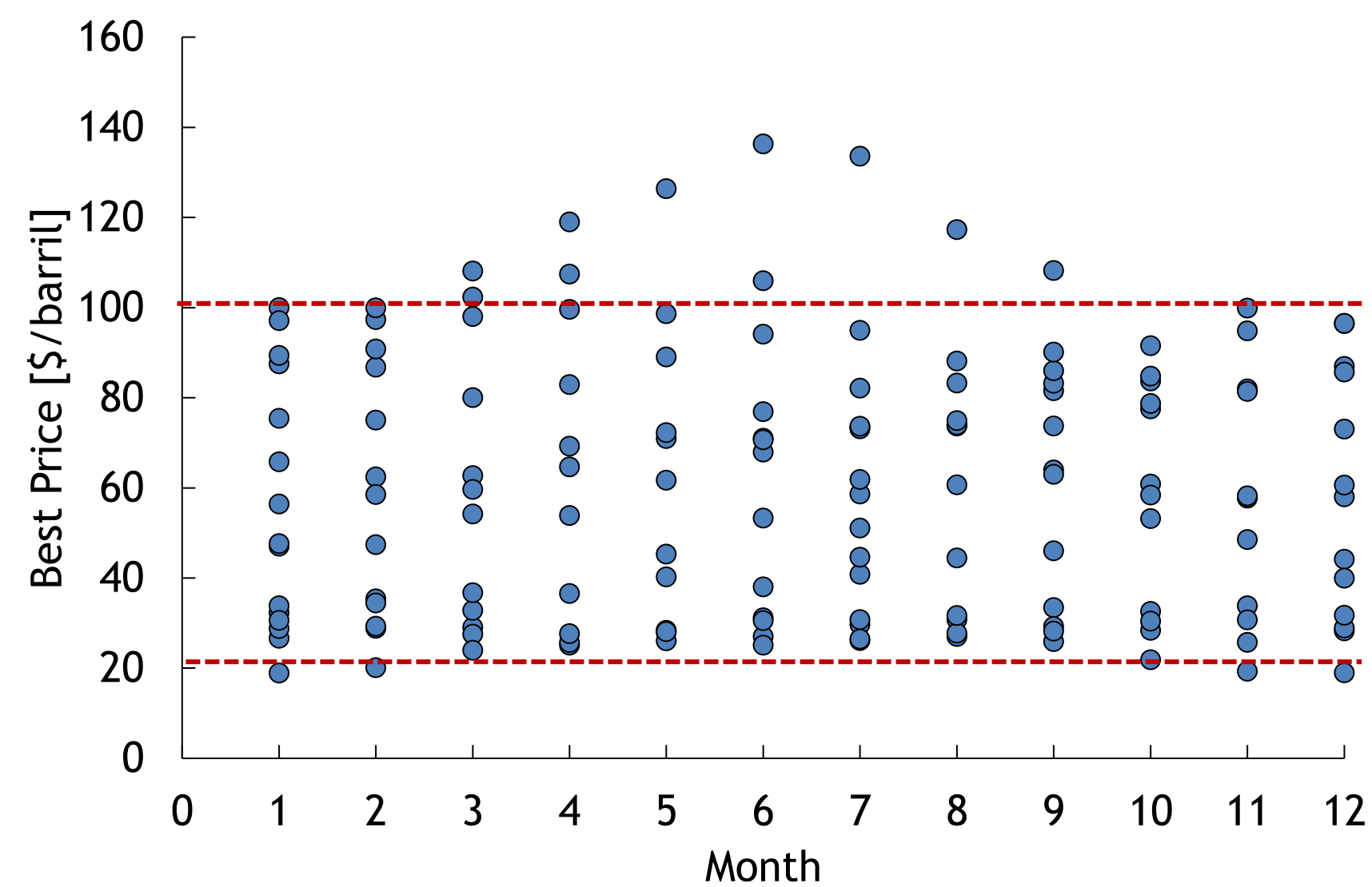
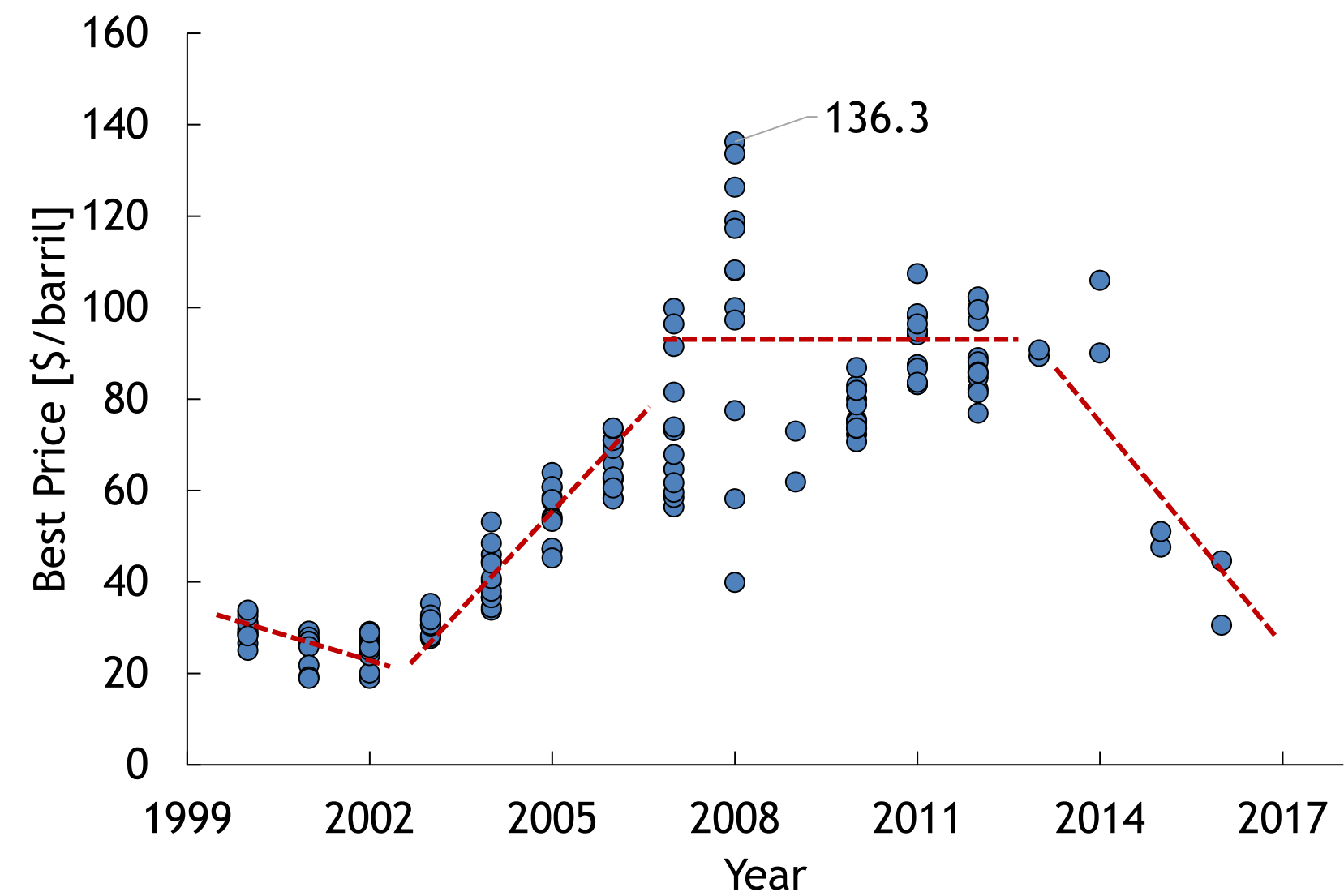
#### a. Externas:

- [https://apmonitor.com/me575/uploads/Main/oil\\_data.txt](https://apmonitor.com/me575/uploads/Main/oil_data.txt)
- Precio del petróleo (BP).
- WTI
- Precio Henry Hub del gas (HH).
- Precio Mont Belvieu del propano (GNL).

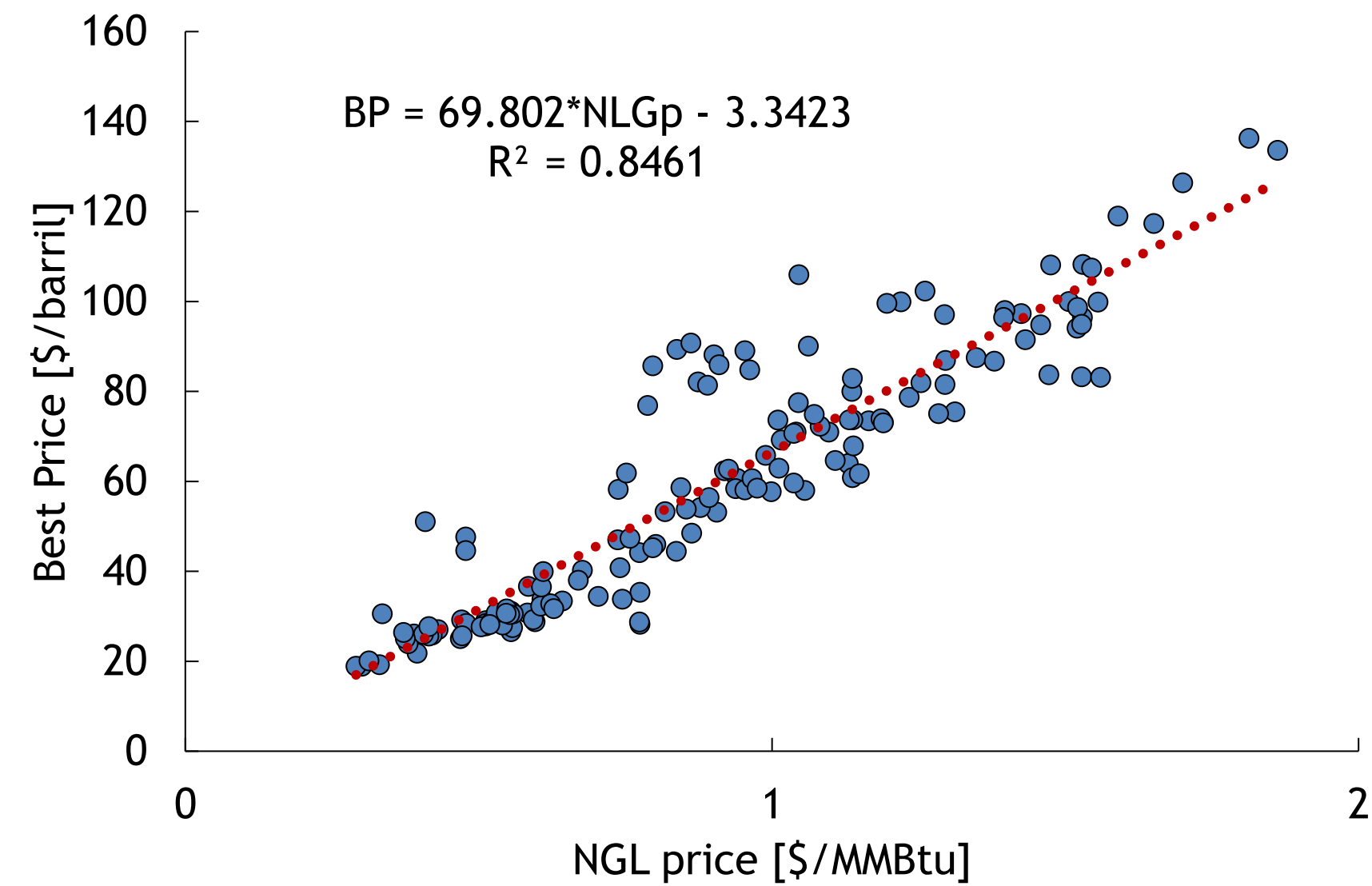
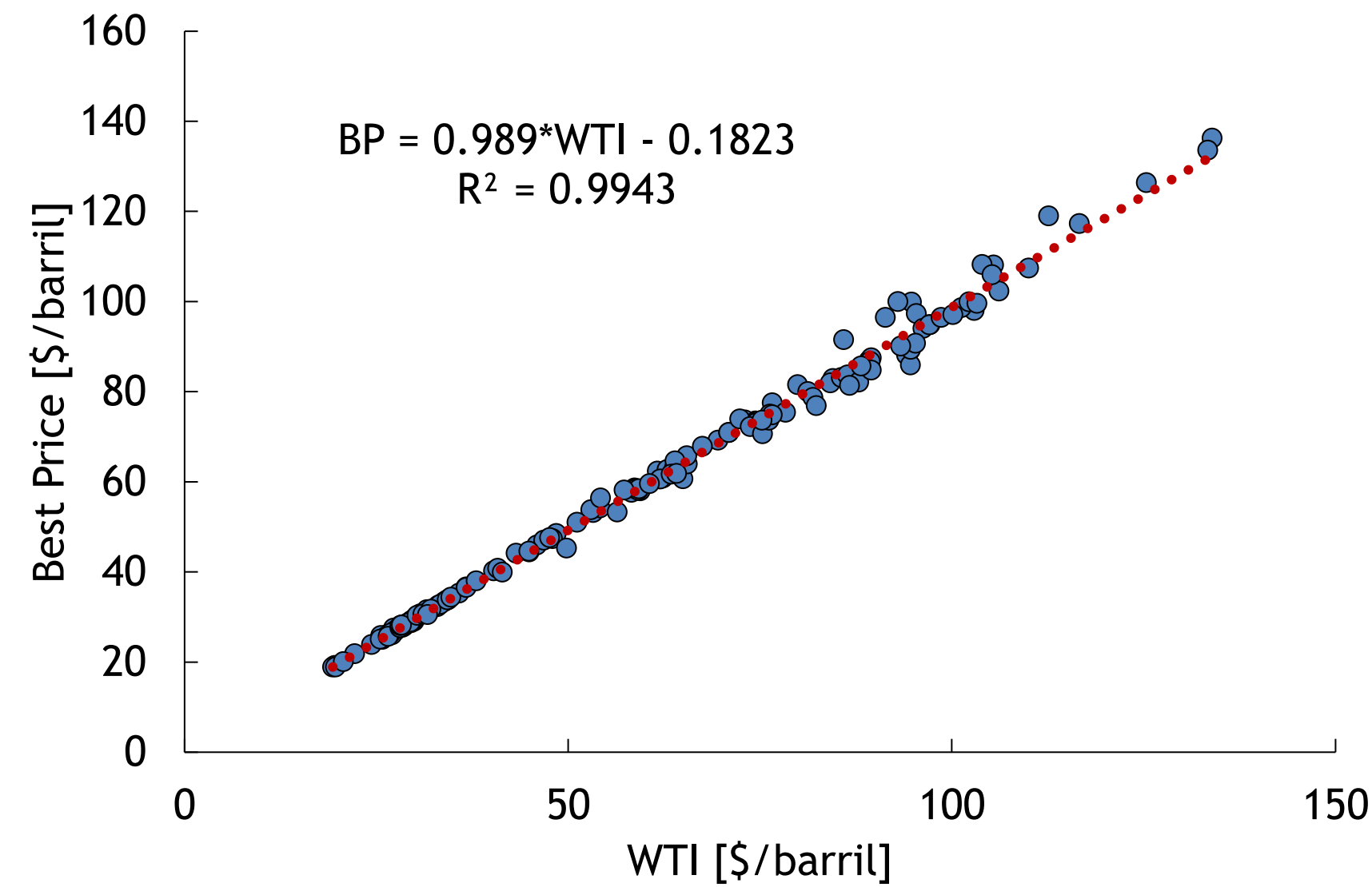
1. Descargar archivo.
2. Abrirlo en Excel.
3. Graficar los datos.
4. Determinar si existen relaciones.
5. Proponer modelos individuales.
6. Proponer modelo global para  $BP(WTI, HH, GNL)$

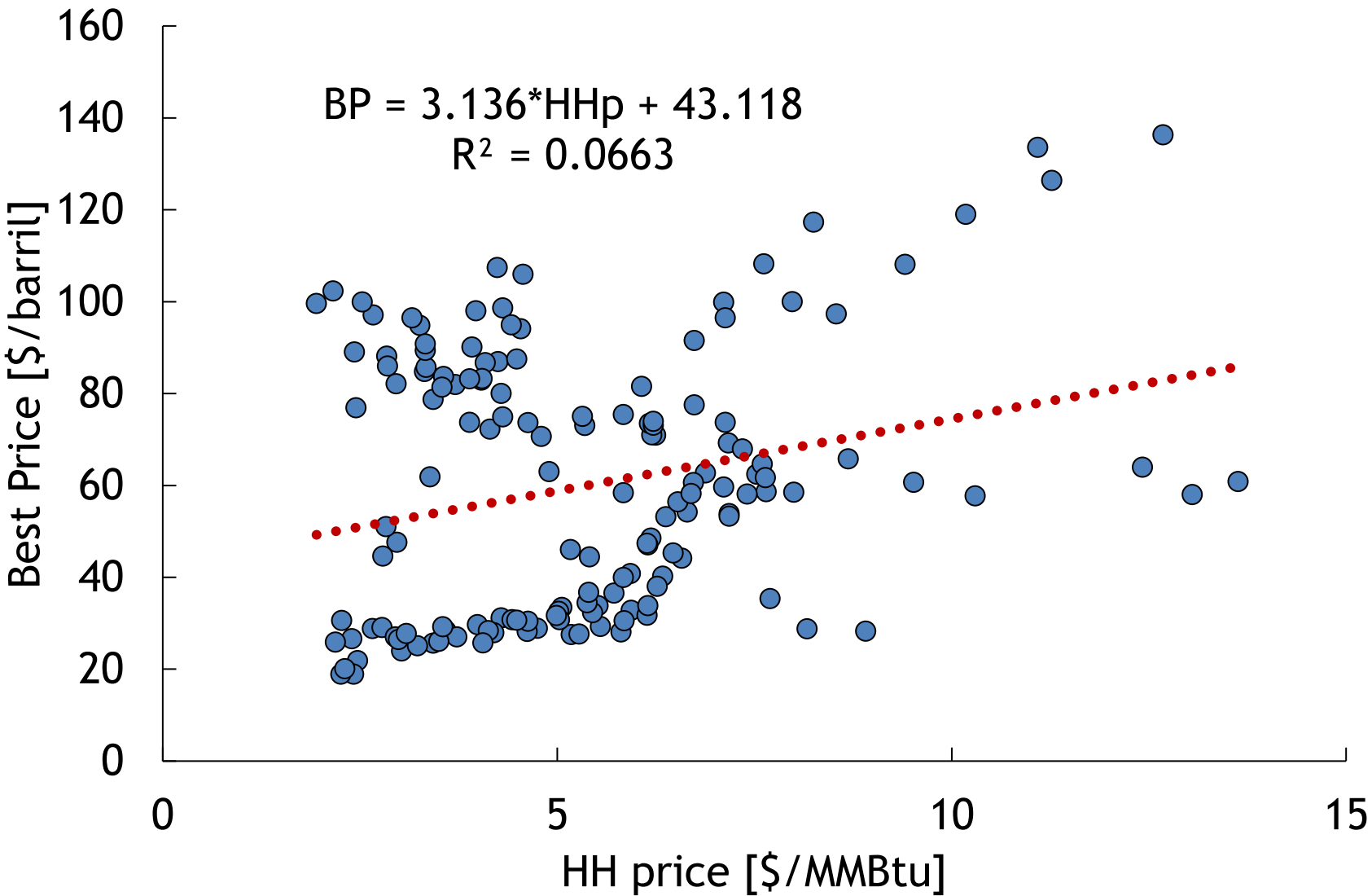


MANEJO DE DATOS - EJERCICIO 1



MANEJO DE DATOS - EJERCICIO 1





¿Modelo global se propone para  
BP(WTI, HH, GNL)?

### Fuentes de datos:

#### a. Externas:

- <https://regressit.com>
- Consumo de gas natural.
- Temperaturas ambientales mínimas, máximas y promedio.

1. Descargar archivo:
  - NC\_natural\_gas\_consumption\_analysis.xlsx
2. Explorar los datos.
3. Identificar los datos de interés.
  - Fechas.
  - Consumo mensual de gas natural.
  - Temperaturas máximas, mínimas y promedio.
4. Separar los datos de interés.

### Fuentes de datos:

#### a. Externas:

- <https://regressit.com>
- Consumo de gas natural.
- Temperaturas ambientales mínimas, máximas y promedio.

#### 5. Filtrar datos:

- Identificar rangos e inconsistencias.

#### 6. Buscar duplicados:

- Función CONTAR.SI

#### 7. Graficar todos los datos e identificar tendencias.

#### 8. Obtener coeficientes de correlación (grado de variación conjunta entre dos variables):

- Valor de 1 ó -1: correlación fuerte.
- Valor de 0: no hay correlación.



### Fuentes de datos:

#### a. Externas:

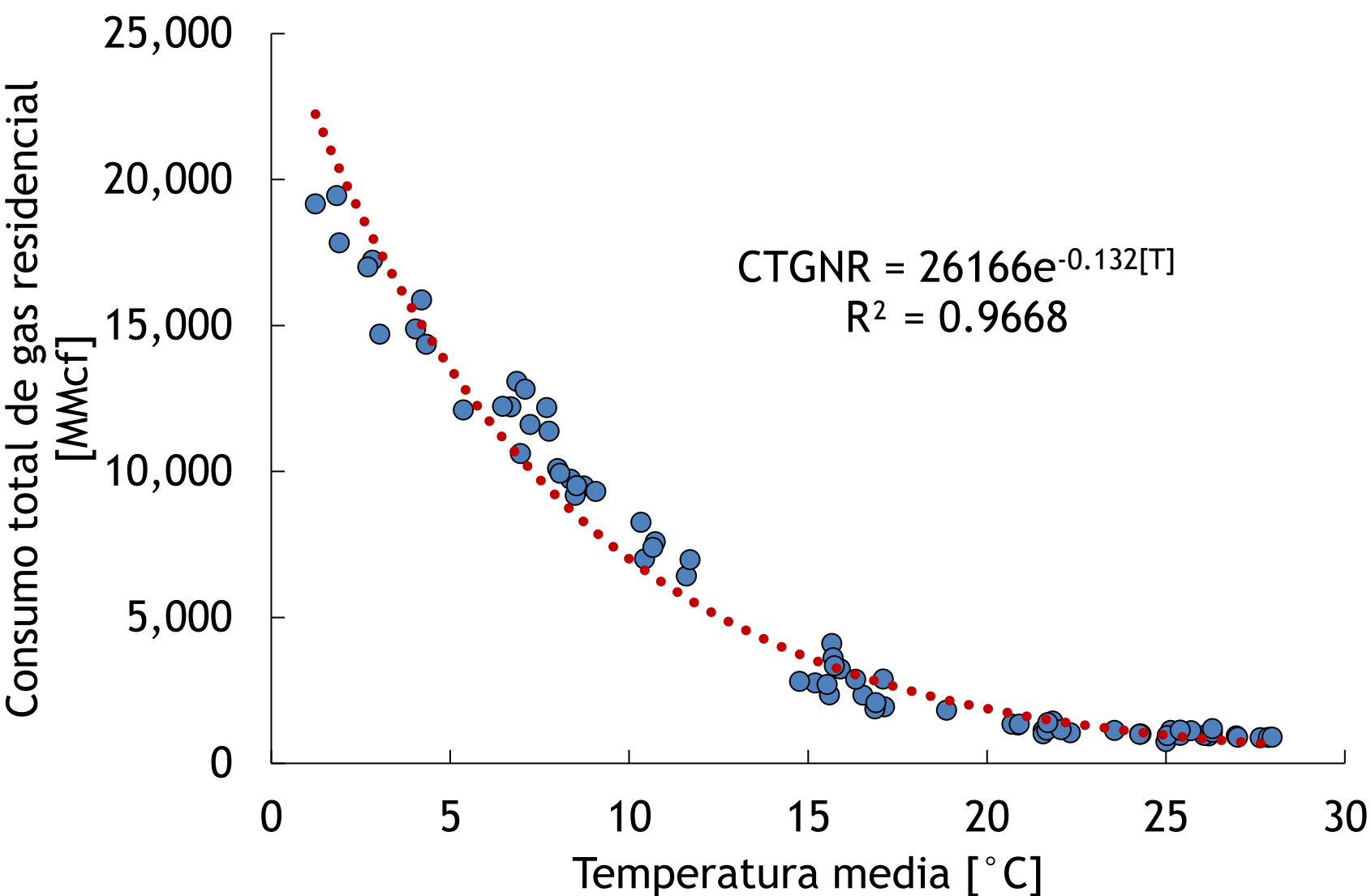
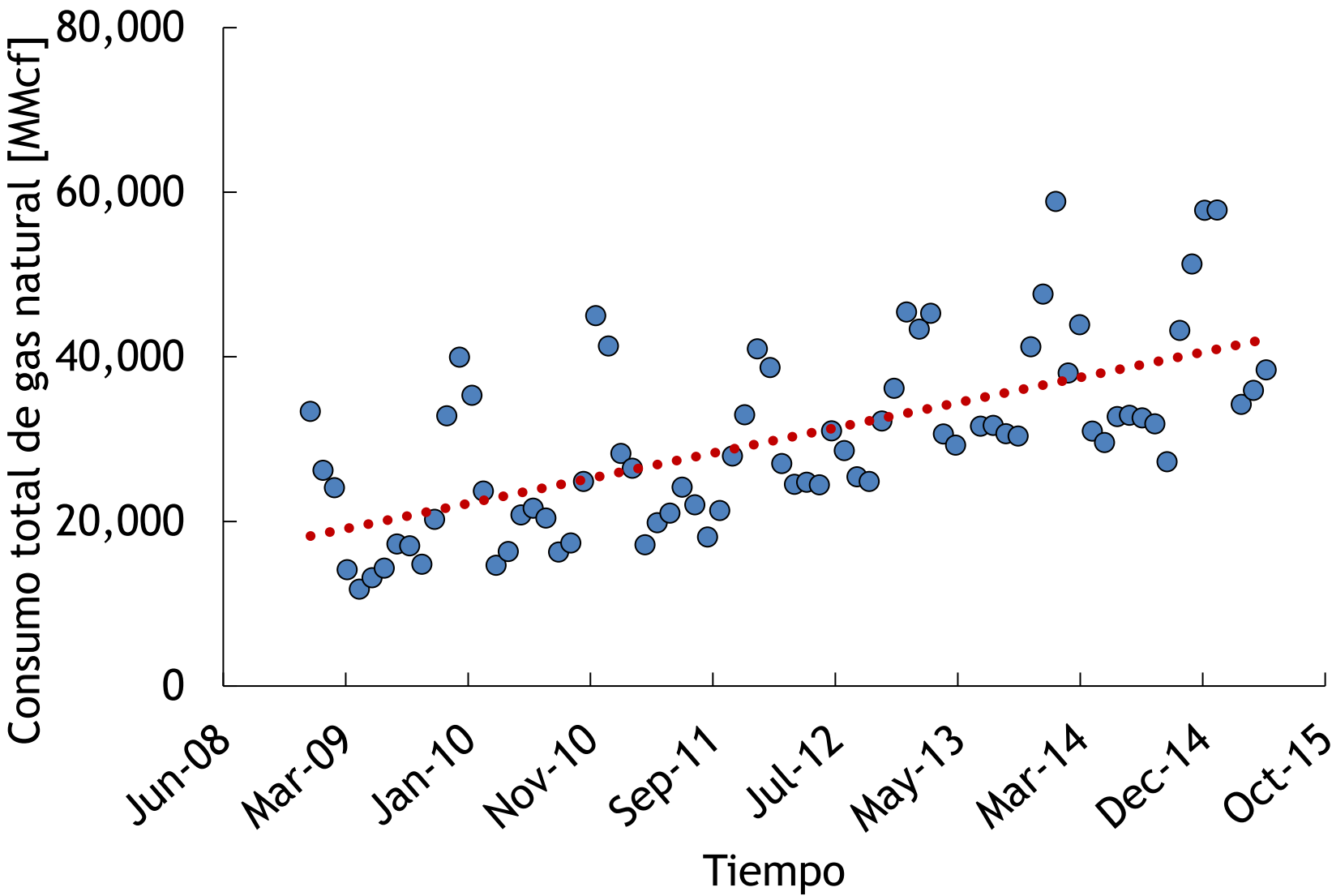
- <https://regressit.com>
- Consumo de gas natural.
- Temperaturas ambientales mínimas, máximas y promedio.

#### 9. Tablas dinámicas:

- Promedios y el total de una variable.
- Gráficos dinámicos.

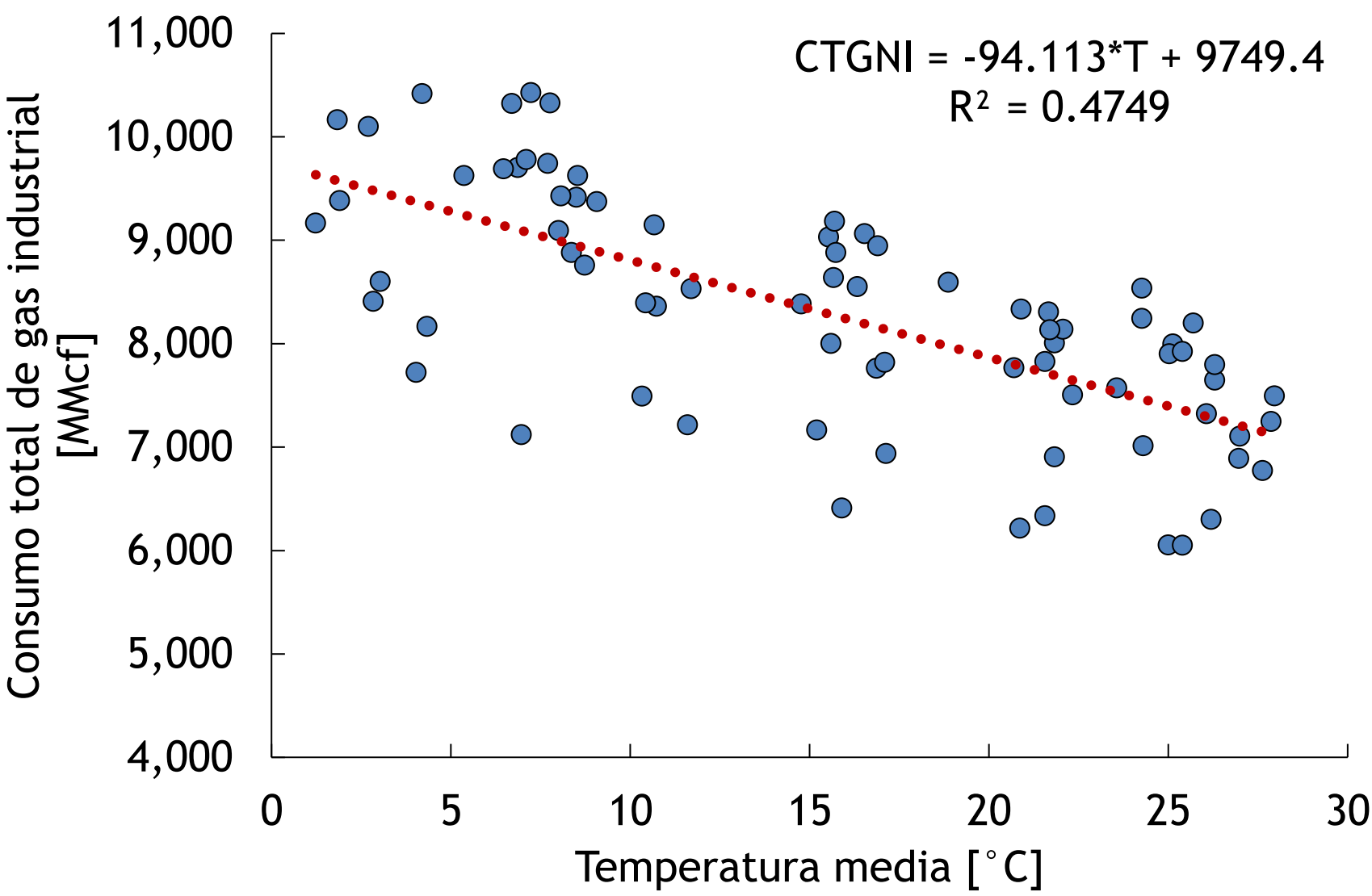
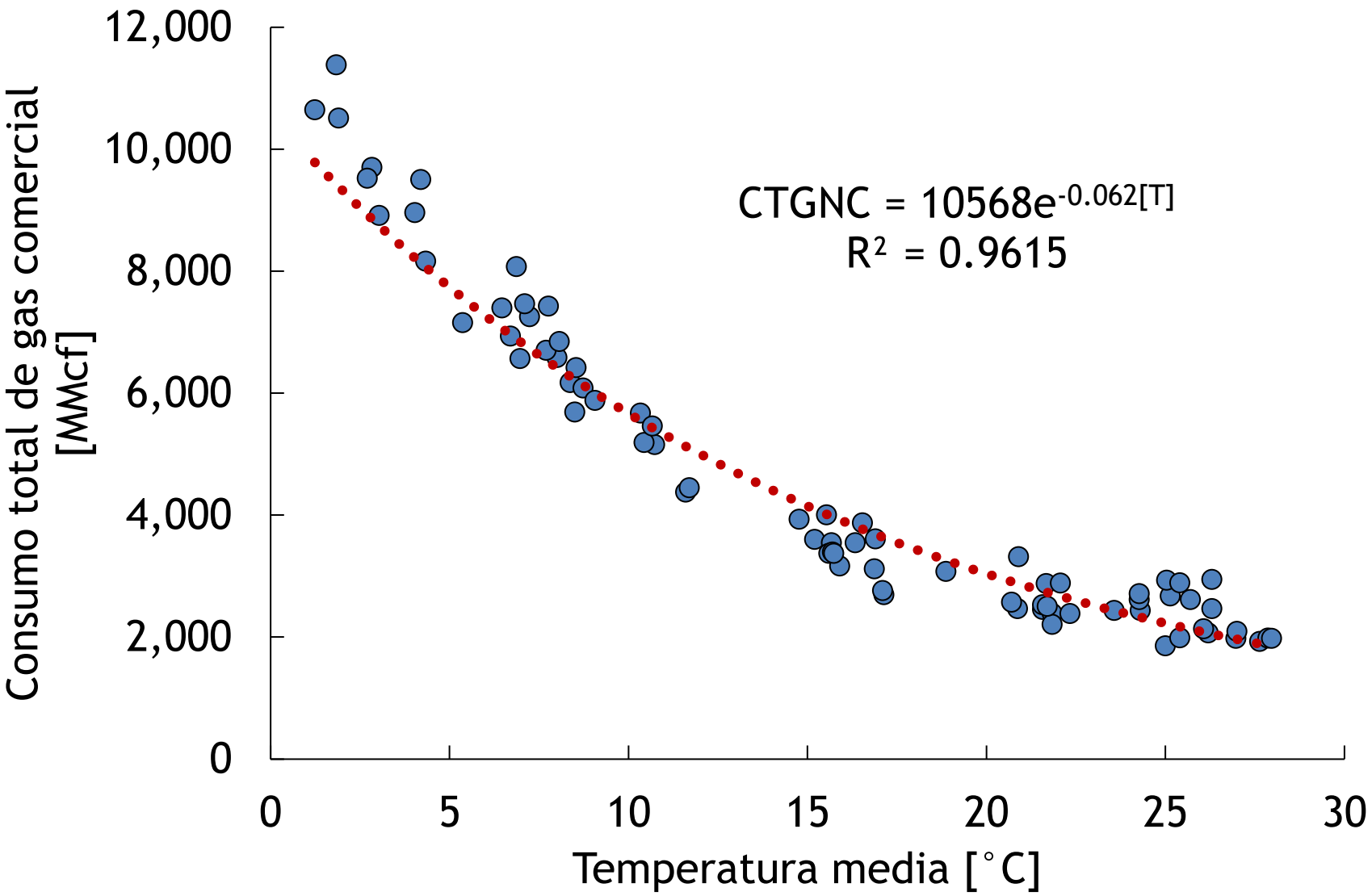
10. Analizar gráficamente cada tipo de consumo de gas natural *versus* la temperatura media.

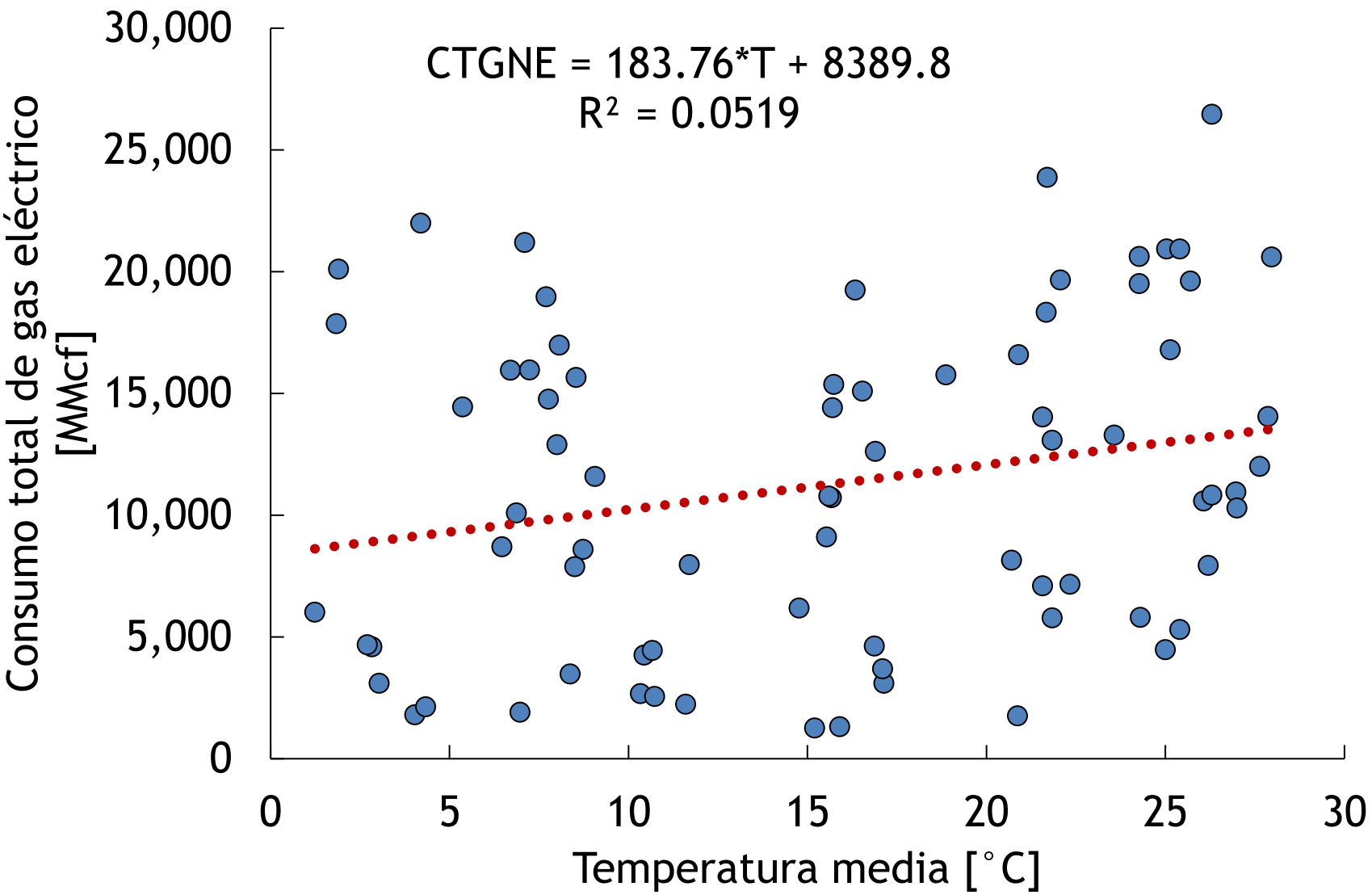
MANEJO DE DATOS - EJERCICIO 2





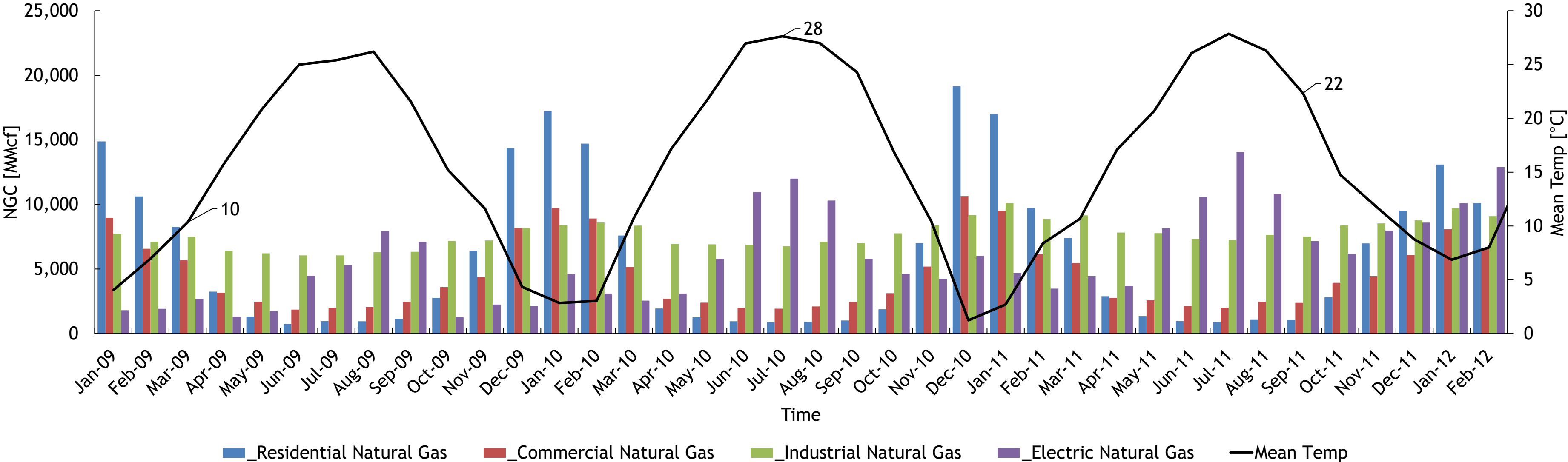
MANEJO DE DATOS - EJERCICIO 2





¿Modelo global se propone para el consumo de gas natural en función de la temperatura?

Gráfica del consumo de gas natural por sector y temperatura media



### 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](https://www.weather.gov/key/climate_heat_cool)

### Fuentes de datos:

#### a. Externas:

- [https://www.weather.gov/key/climate\\_heat\\_cool](https://www.weather.gov/key/climate_heat_cool)
- [https://databank.worldbank.org/source/environment-social-and-governance-\(esg\)-data/Series](https://databank.worldbank.org/source/environment-social-and-governance-(esg)-data/Series)

### EJERCICIO 3

#### Fuentes de datos:

##### a. Externas:

- <https://regressit.com>
- Consumo de gas natural por sector.
- Temperaturas ambientales.
- Parámetros de consumo de energía.

1. Realizar el procedimiento descrito anteriormente con el archivo proporcionado.
2. Elabore una hoja nueva adicional que considere las variables de interés ya establecidas pero que también contenga las siguientes:
  - Cooling Degree Days.
  - Heating Degree Days.

### EJERCICIO 3

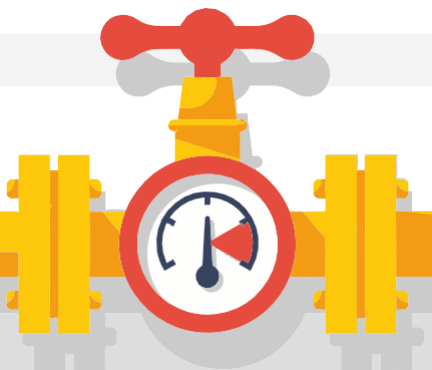
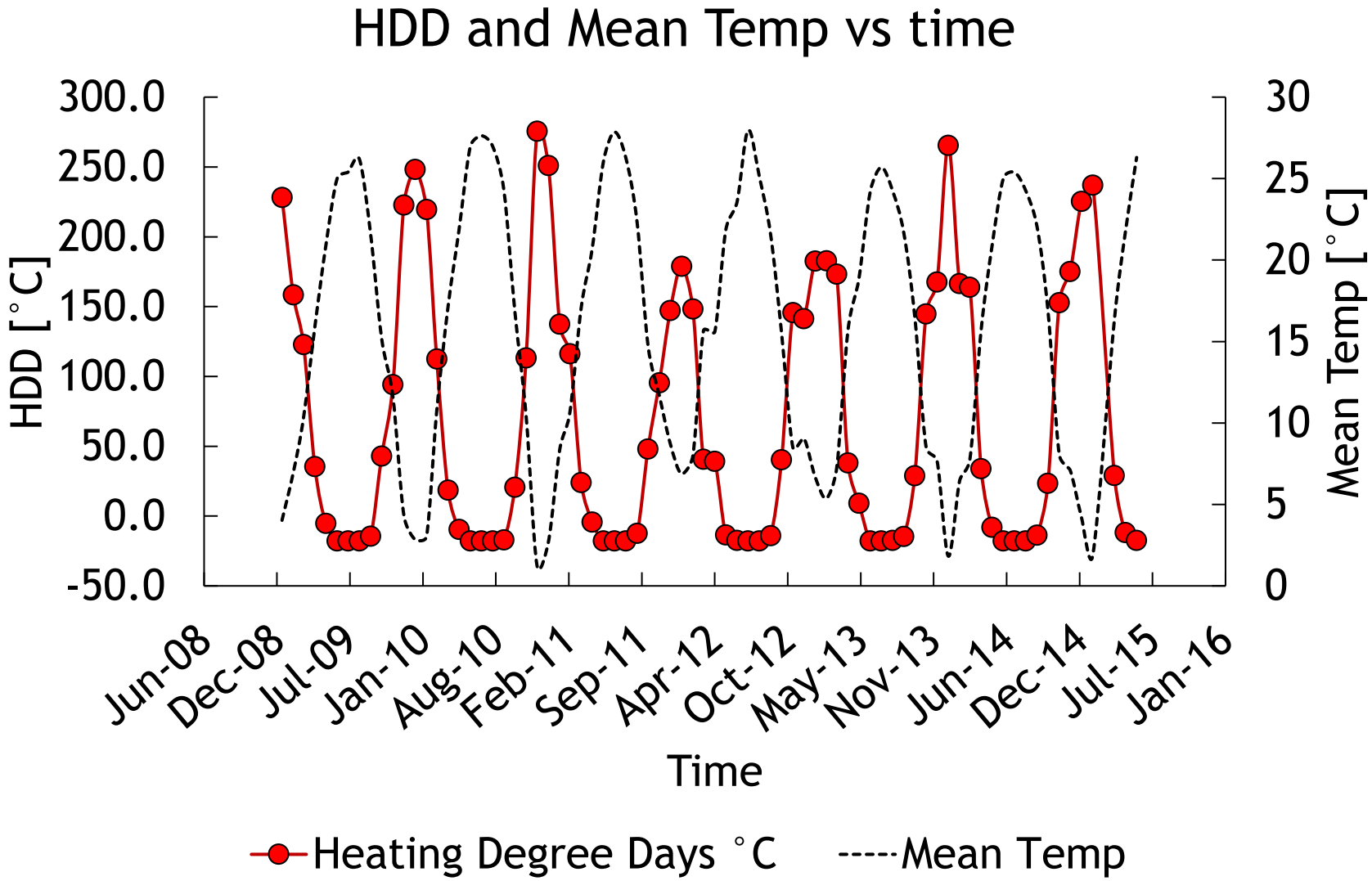
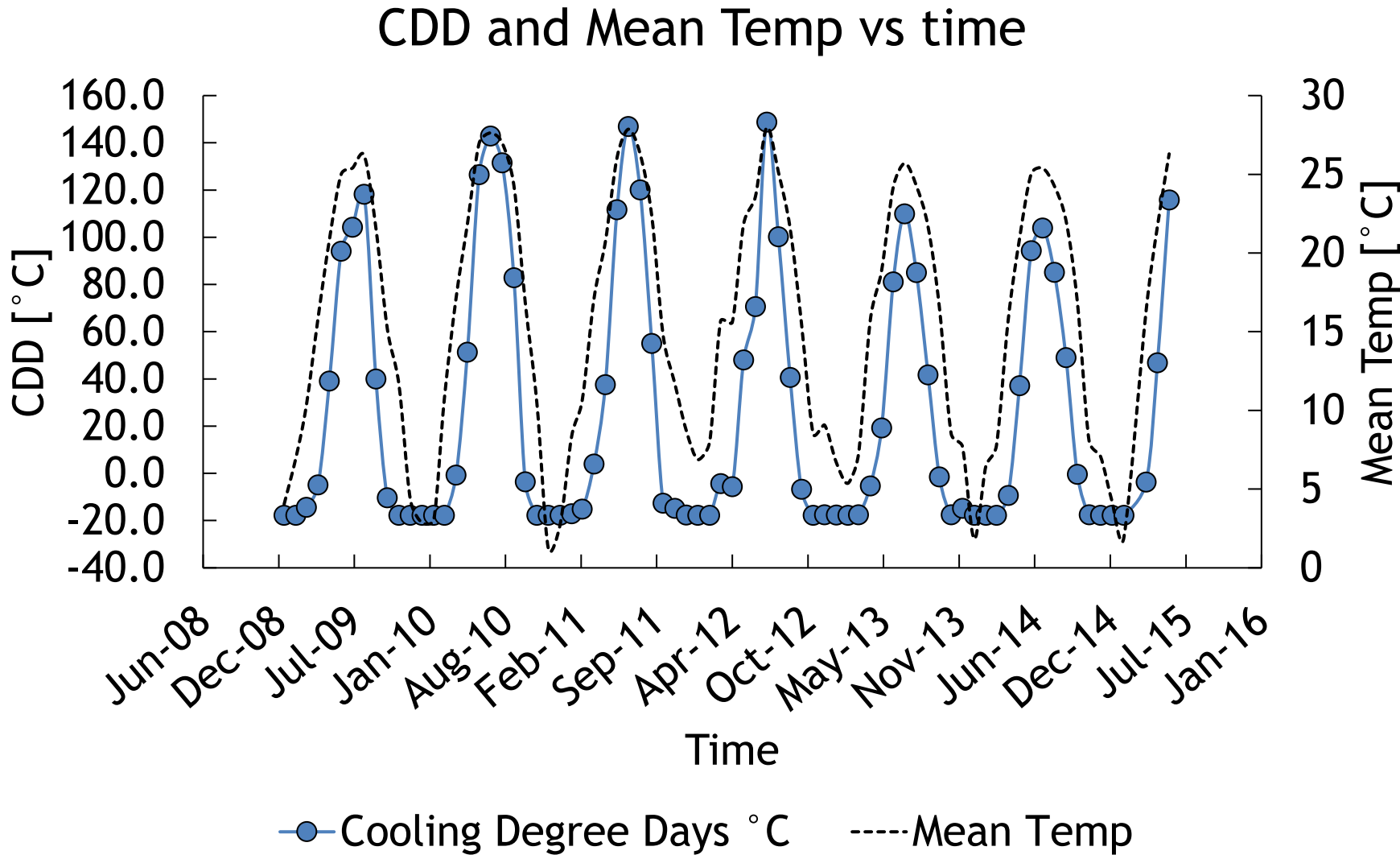
#### Fuentes de datos:

##### a. Externas:

- <https://regressit.com>
- Consumo de gas natural por sector.
- Temperaturas ambientales.
- Parámetros de consumo de energía.

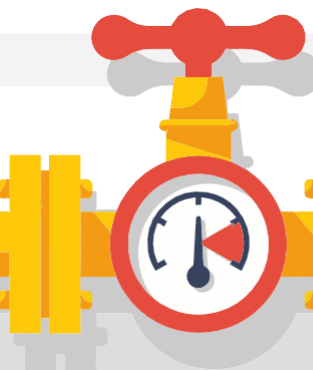
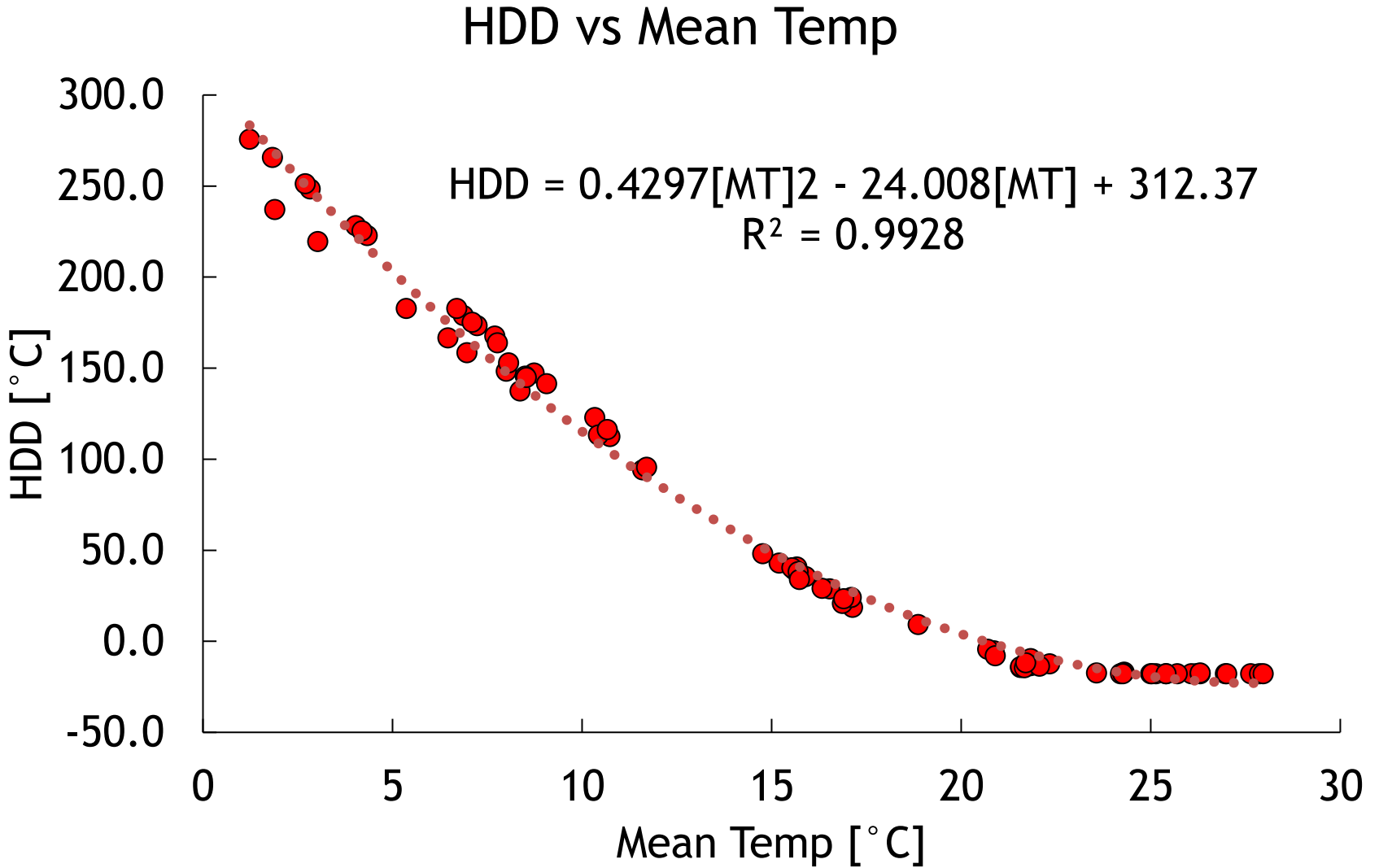
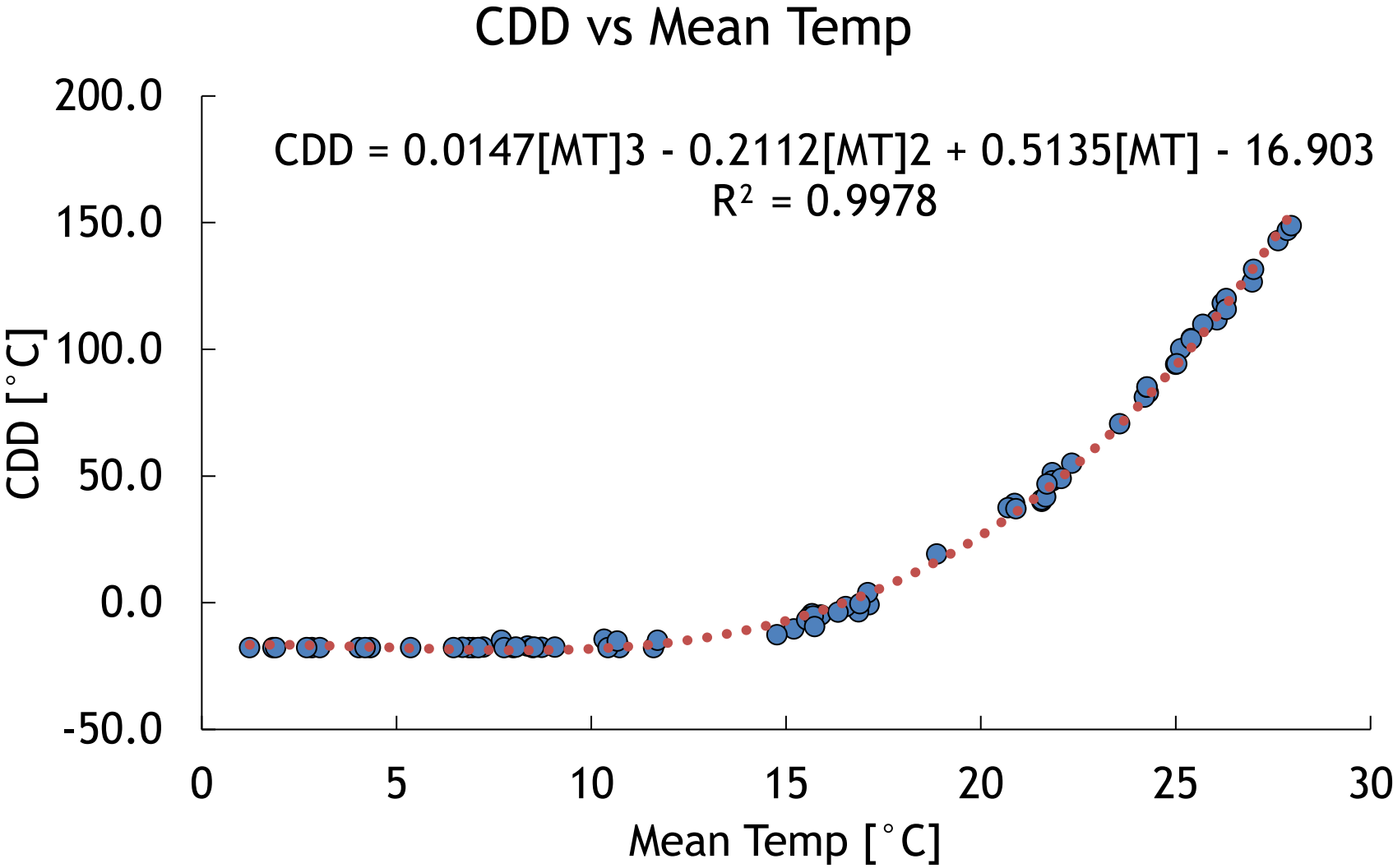
3. Conserve el archivo para la elaboración de las próximas regresiones.
4. Envíe por correo su archivo elaborado.

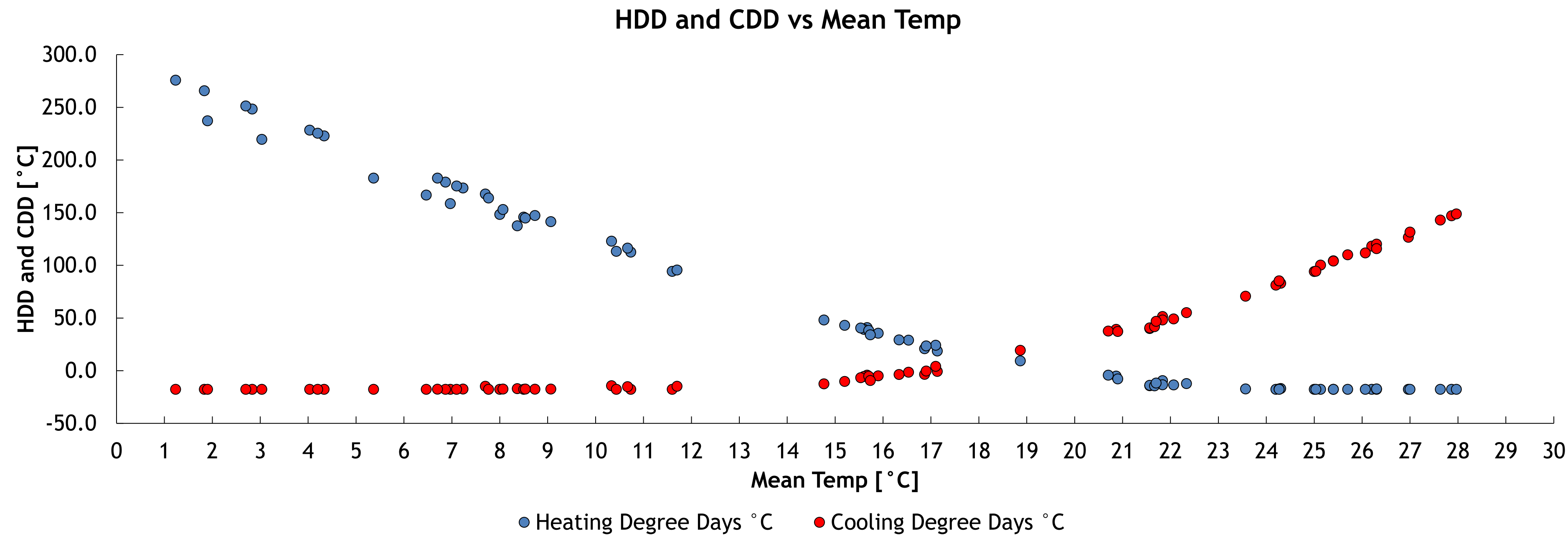






MANEJO DE DATOS - EJERCICIO 3





### Fuentes de datos:

#### b. Internas:

- Datos de operación.
- Registros de pozos.
- Datos de laboratorio.

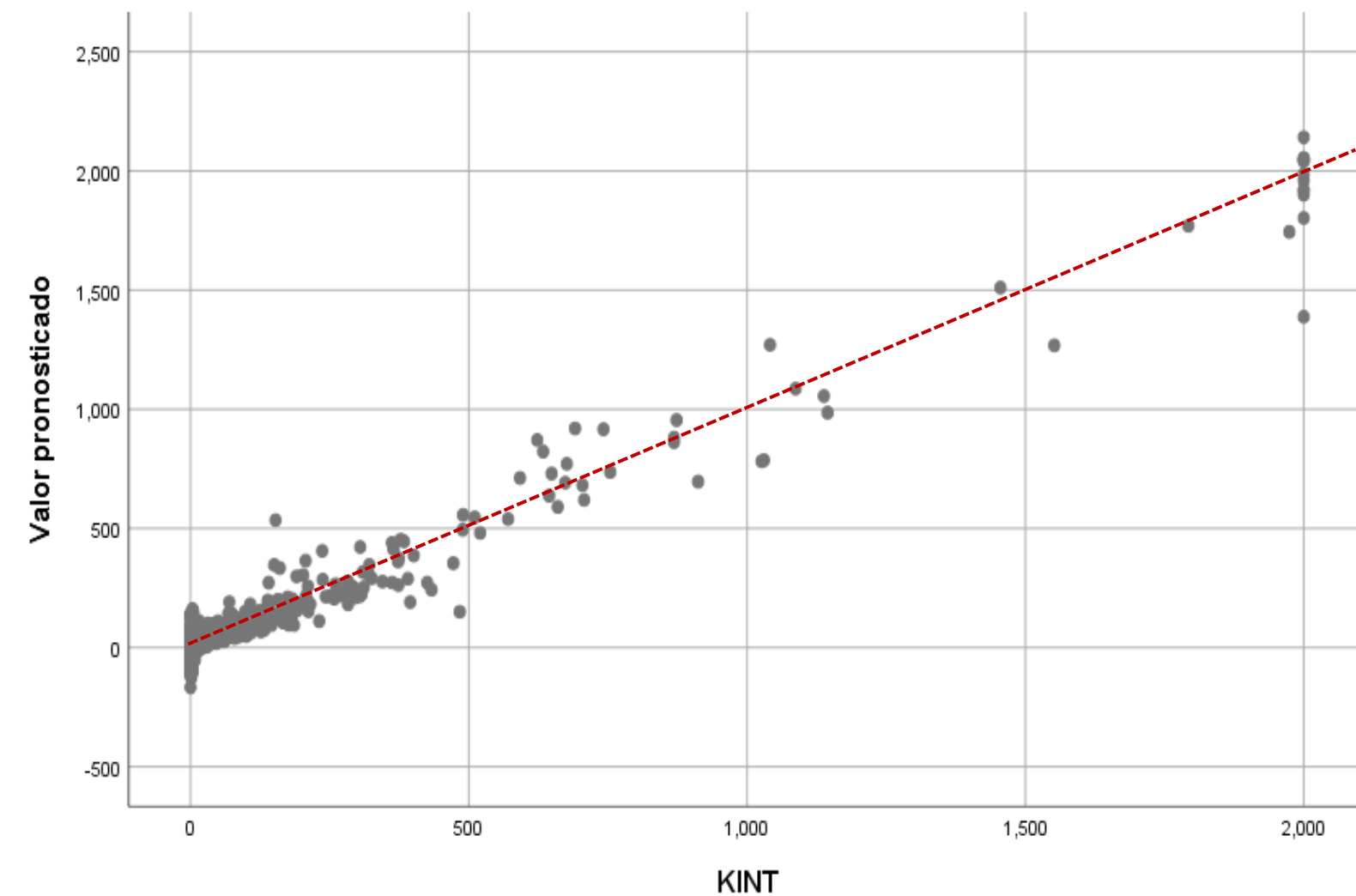


### Fuentes de datos:

#### b. Internas:

- Registros de pozos:
- Estimación de la permeabilidad en función de 18 variables.

### Resultados de aplicación de una red neuronal (RN) a datos petrofísicos

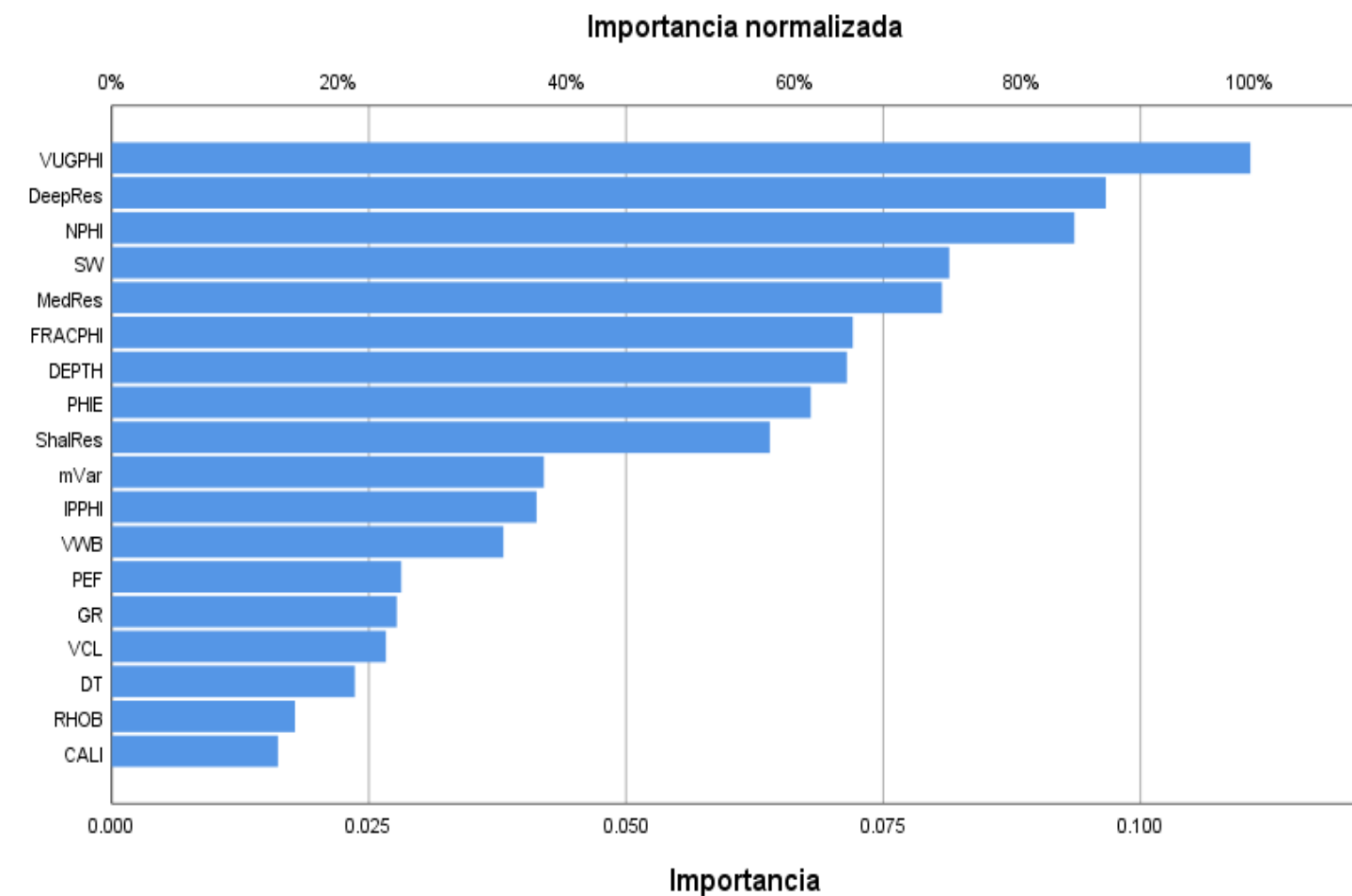


## Fuentes de datos:

### b. Internas:

- Registros de pozos:
- Variables con más influencia:
  - Garganta de poro media respecto a porosidad (VUGPHI).
  - Resistividad profunda (DeepRes).
  - Porosidad neutrón (NPHI).

Gráfico de importancia normalizada de cada parámetro utilizado en la RN



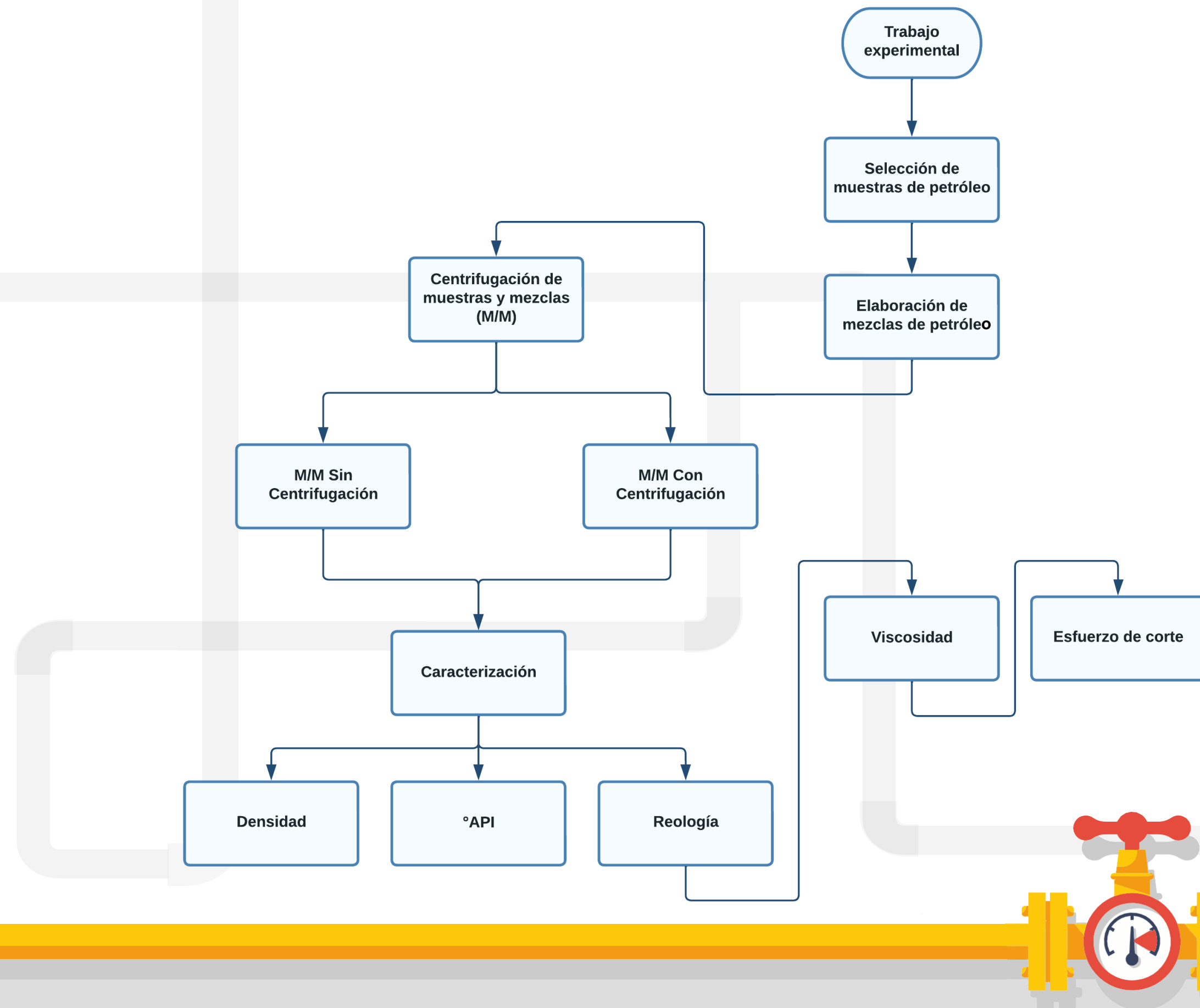
## Fuentes de datos:

### b. Internas:

- Datos de laboratorio.

Objetivo: Estimación de viscosidad.

Trabajo experimental.



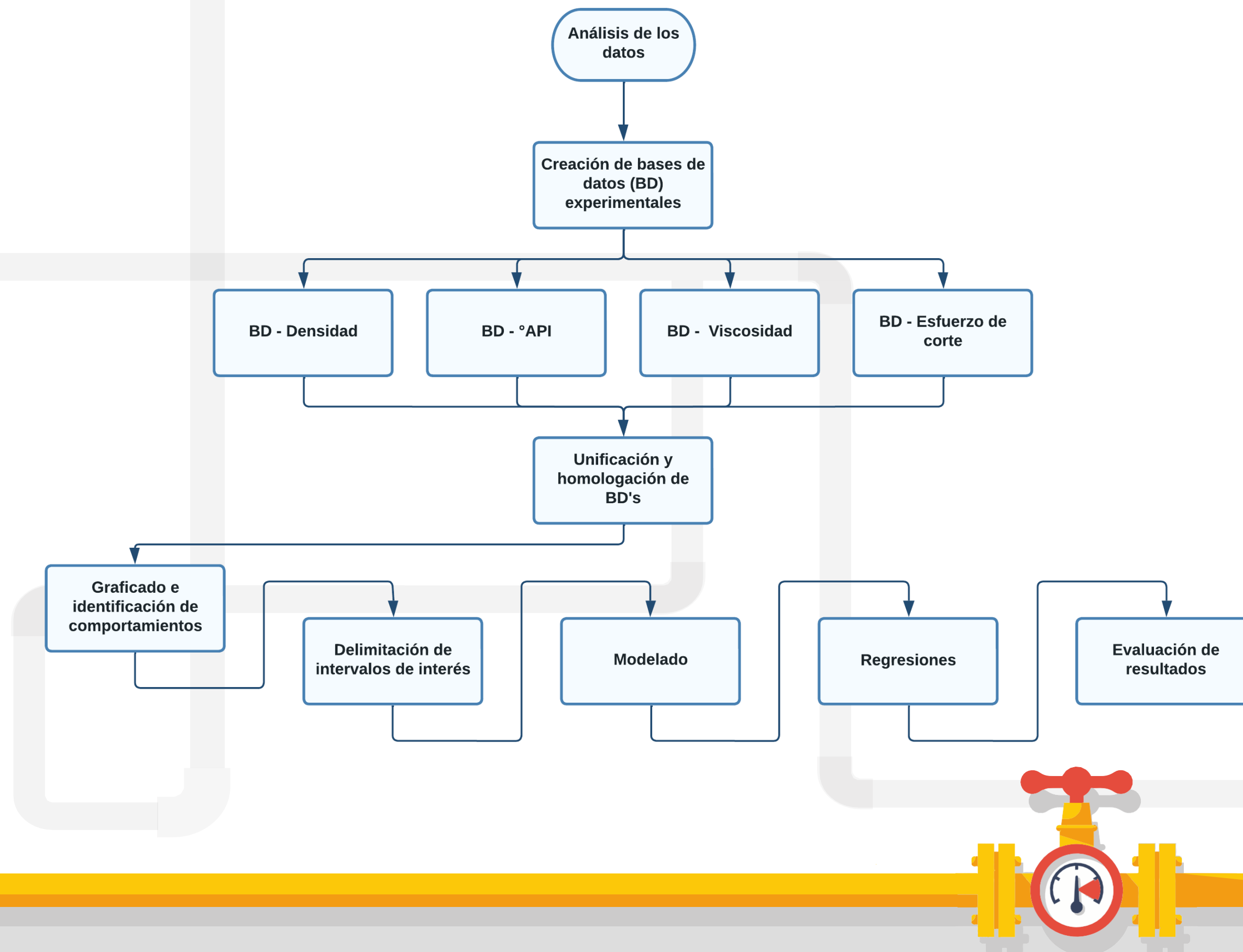
## Fuentes de datos:

### b. Internas:

- Datos de laboratorio.

Objetivo: Estimación de viscosidad.

Trabajo de análisis de datos.

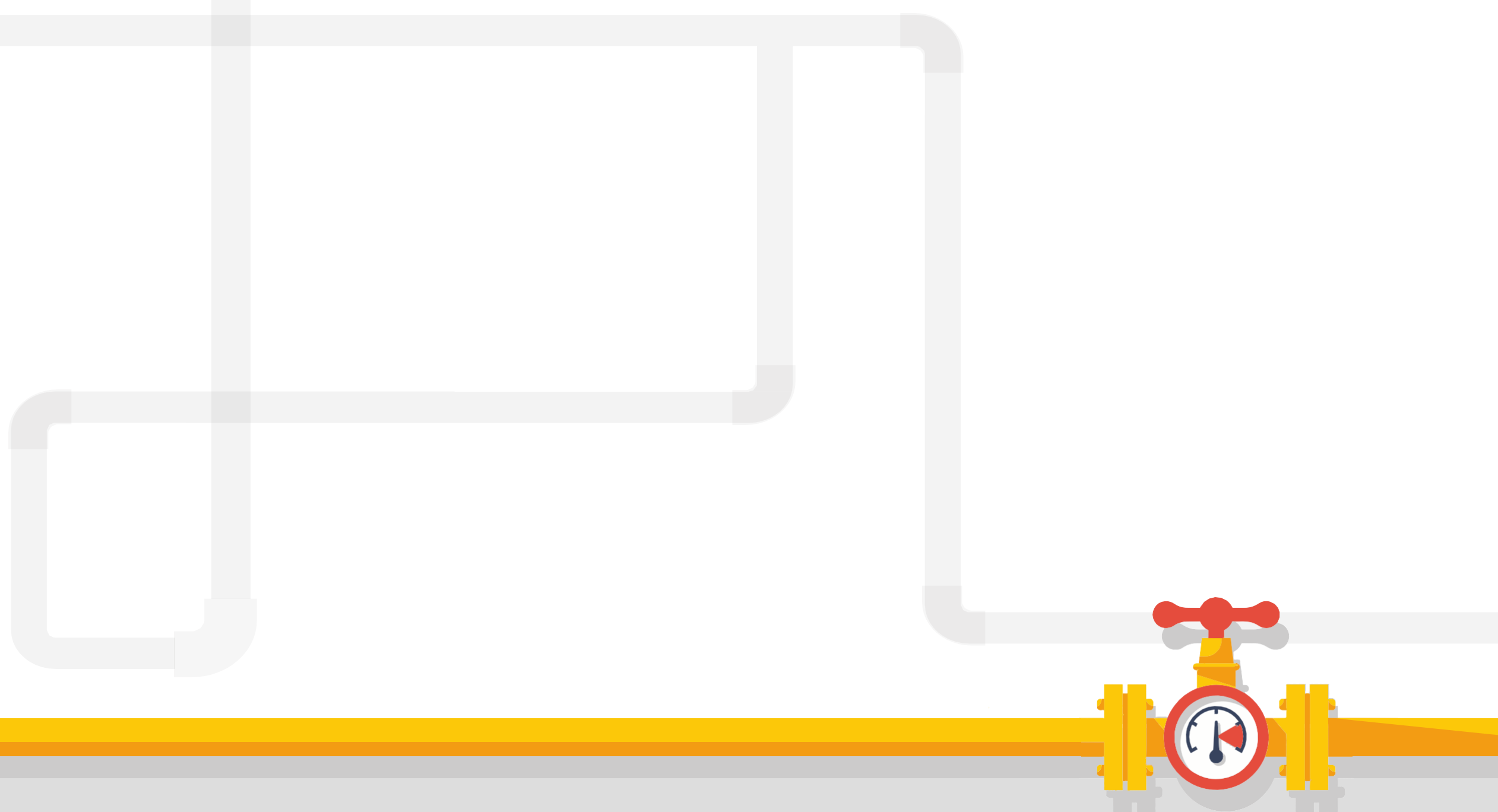


# EJERCICIOS DE REFORZAMIENTO

---

Quiz #3:

<https://acortar.link/Ok7IIQ>





Previo a las regresiones:

- Preparar una base de datos propia.
- Selecciona la variable objetivo y las variables independientes.
- Define los rangos de análisis.
- Define el modelo.

Variable Ind. #1	Variable Ind. #2	Variable Ind. #3	Variable Depend.
...	...	...	...
...	...	...	...



Universidad Nacional de Colombia  
SPE Student Chapter

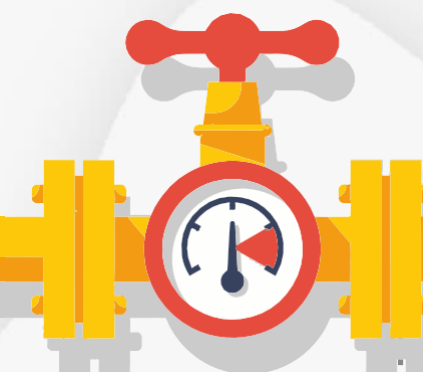
PETROIntelligence.com

## SESIÓN 3 REGRESIONES APLICADAS A LA INGENIERÍA DE PETRÓLEOS (CON PYTHON)

Misael Edgar Zepeda Díaz

[mzepeda@petrointelligence.com](mailto:mzepeda@petrointelligence.com)  
(+52 55 1350 7432)

Agosto 2023



Título	Horas
1. Introducción -----	1
2. Manejo de datos -----	1
3. Python -----	2
4. Regresiones -----	1
5. Análisis de los resultados -----	1
Total =	6

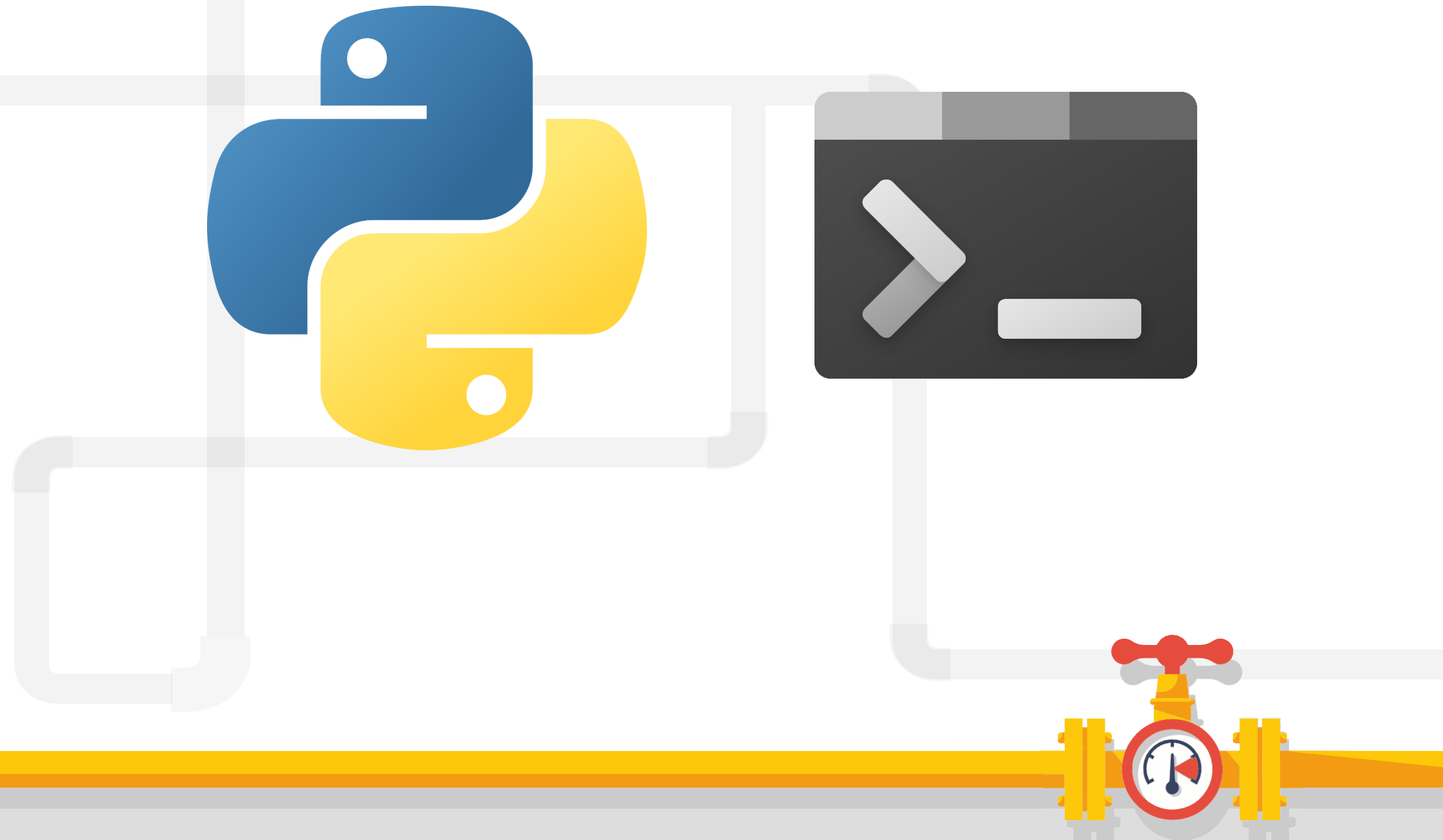
## Python:

Lenguaje de código abierto.

Utilizado en:

1. Aplicaciones web.
2. Desarrollo de software.
3. Ciencia de datos.
4. Machine learning.

Variedad de librerías disponibles.



## Operadores aritméticos:

`+, -, *, /, **`

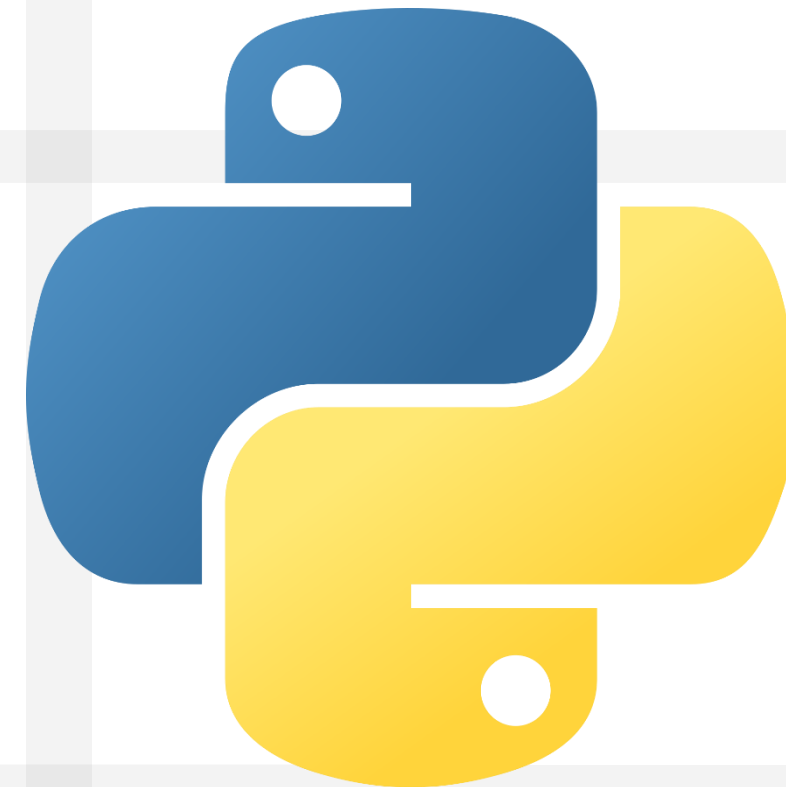
## Comentar código:

`# Esto es un comentario.`

## Imprimir en pantalla:

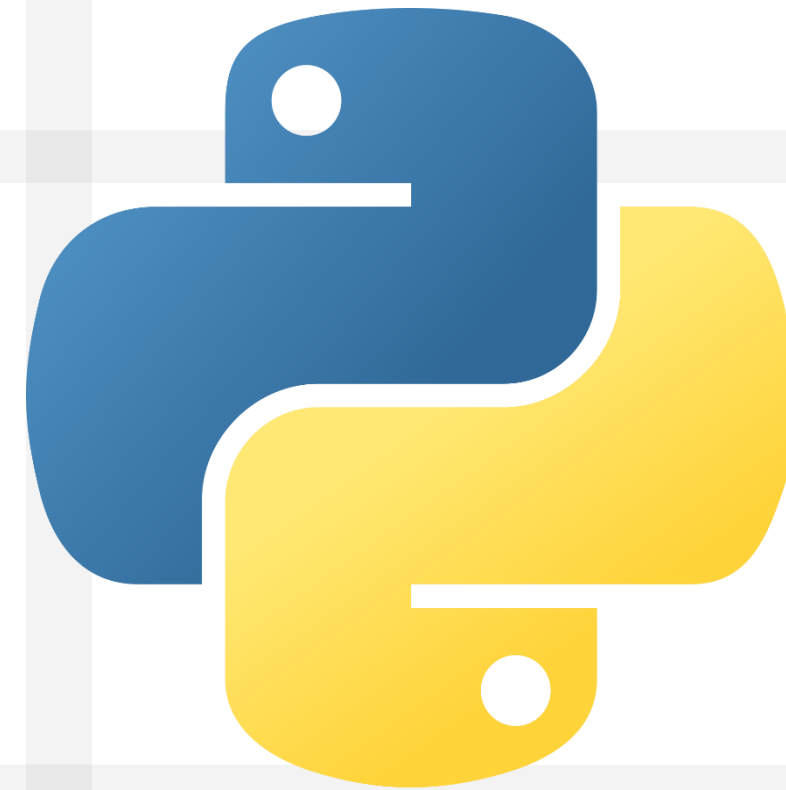
`print("Texto")`

`print(variable)`



## Librerías:

- Pandas - manipulación y análisis de datos.
- Numpy - vectores, matrices y funciones matemáticas.
- Scipy: cómputo científico y algoritmos matemáticos.
- GEKKO: machine learning, optimización y *solvers*.





Editores de texto.



IDE's.





## Google Colab:

Herramienta en la nube que permite escribir y ejecutar código de Python en un navegador.

<https://colab.research.google.com/?hl=es>



## 0\_Ejemplo\_1\_Datos\_CSV

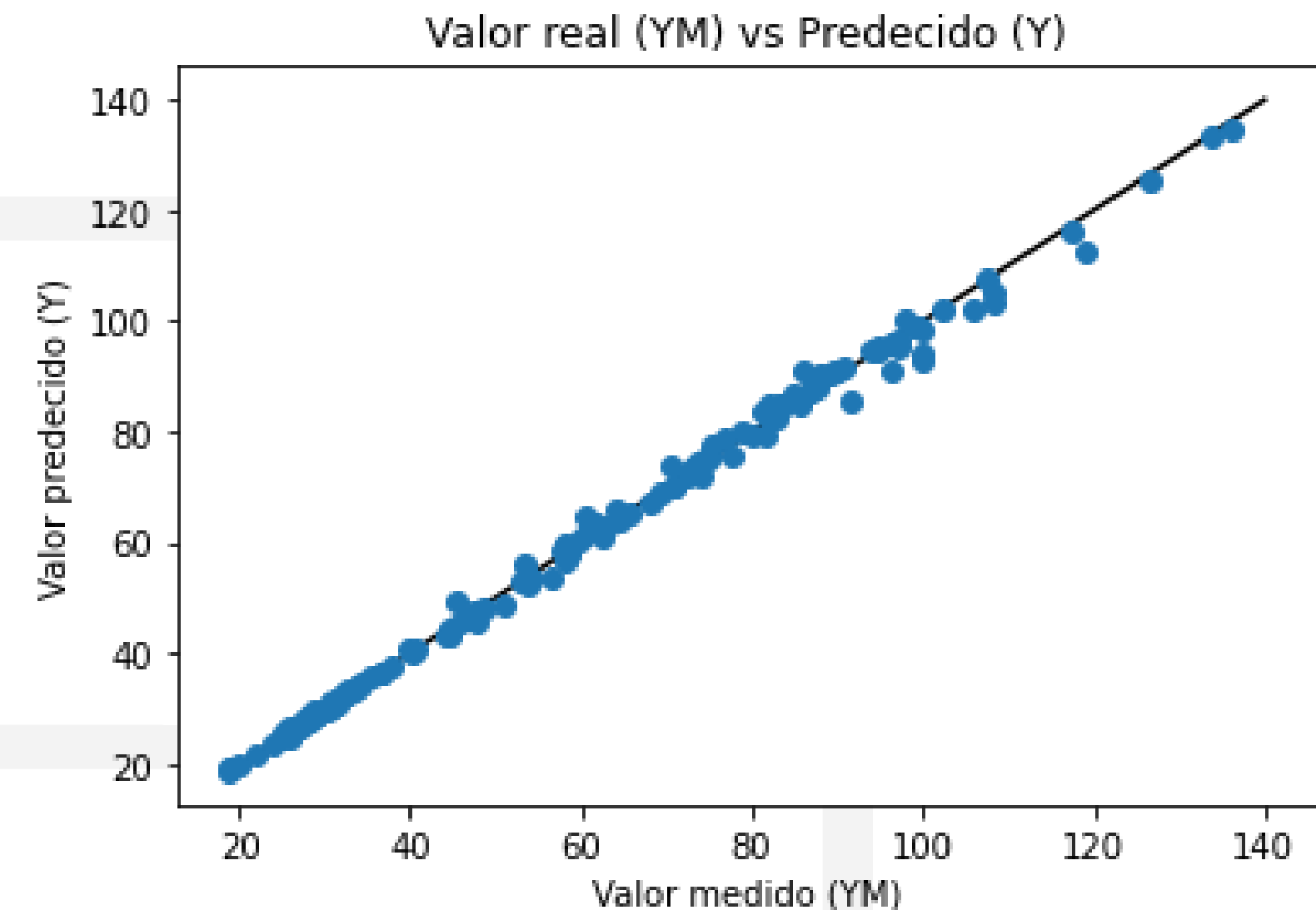
### Objetivos:

1. Explorar la estructura del código.
2. Importar archivo de entrada de datos tipo CSV o desde un TXT.
3. Implementar distintos modelos.

[https://apmonitor.com/me575/uploads/Main/oil\\_data.txt](https://apmonitor.com/me575/uploads/Main/oil_data.txt)

[https://colab.research.google.com/drive/1Sex3hvcTmoF242ZPol8nLlcU2xVv\\_xjN?usp=sharing](https://colab.research.google.com/drive/1Sex3hvcTmoF242ZPol8nLlcU2xVv_xjN?usp=sharing)

$$y = a + b * x_1 + c * x_2 + d * x_3$$

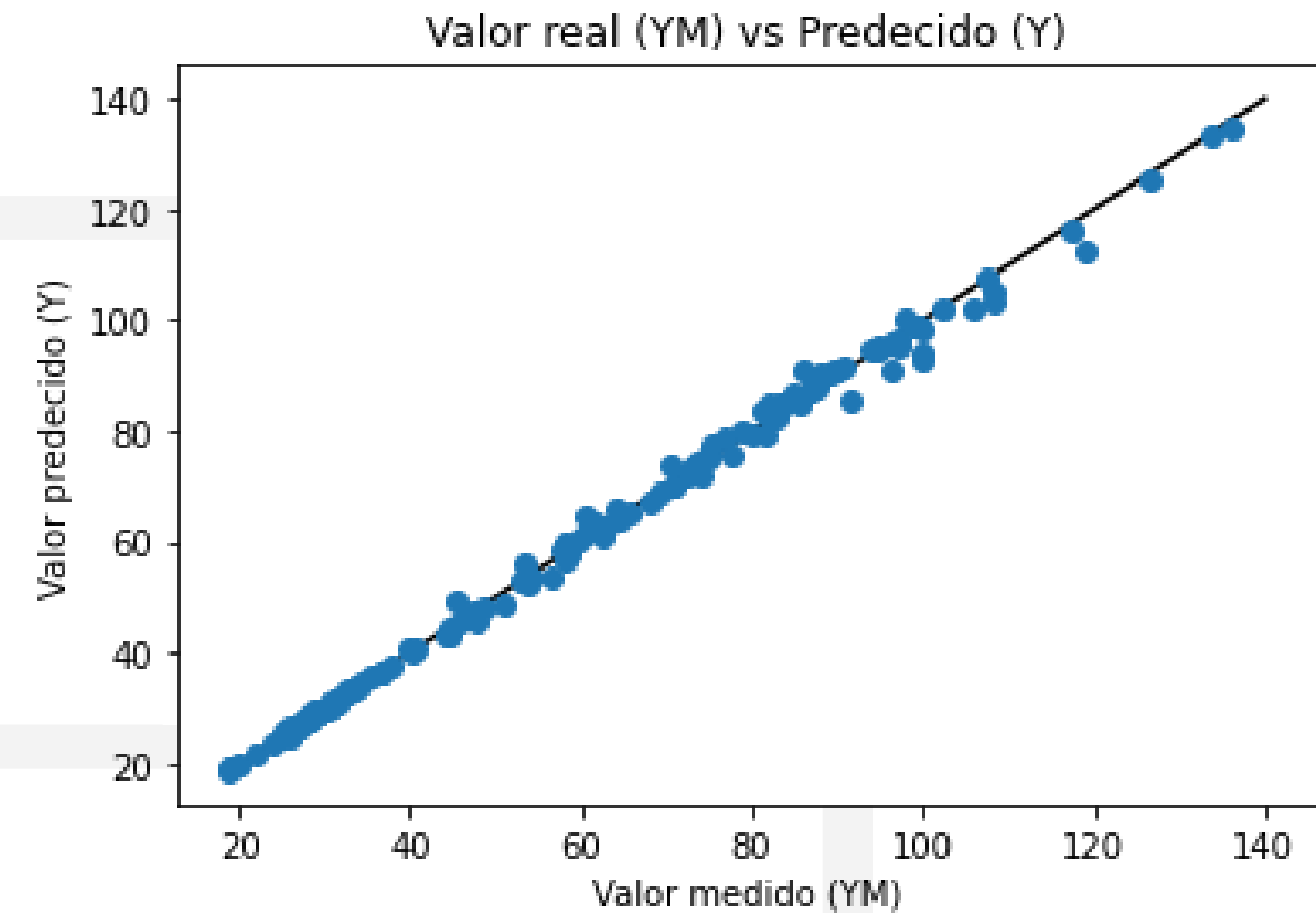


## 0\_Ejemplo\_1\_Datos\_URL

Objetivos:

1. Importar archivo de entrada de datos desde URL.

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



1\_Ejemplo\_1\_CSV

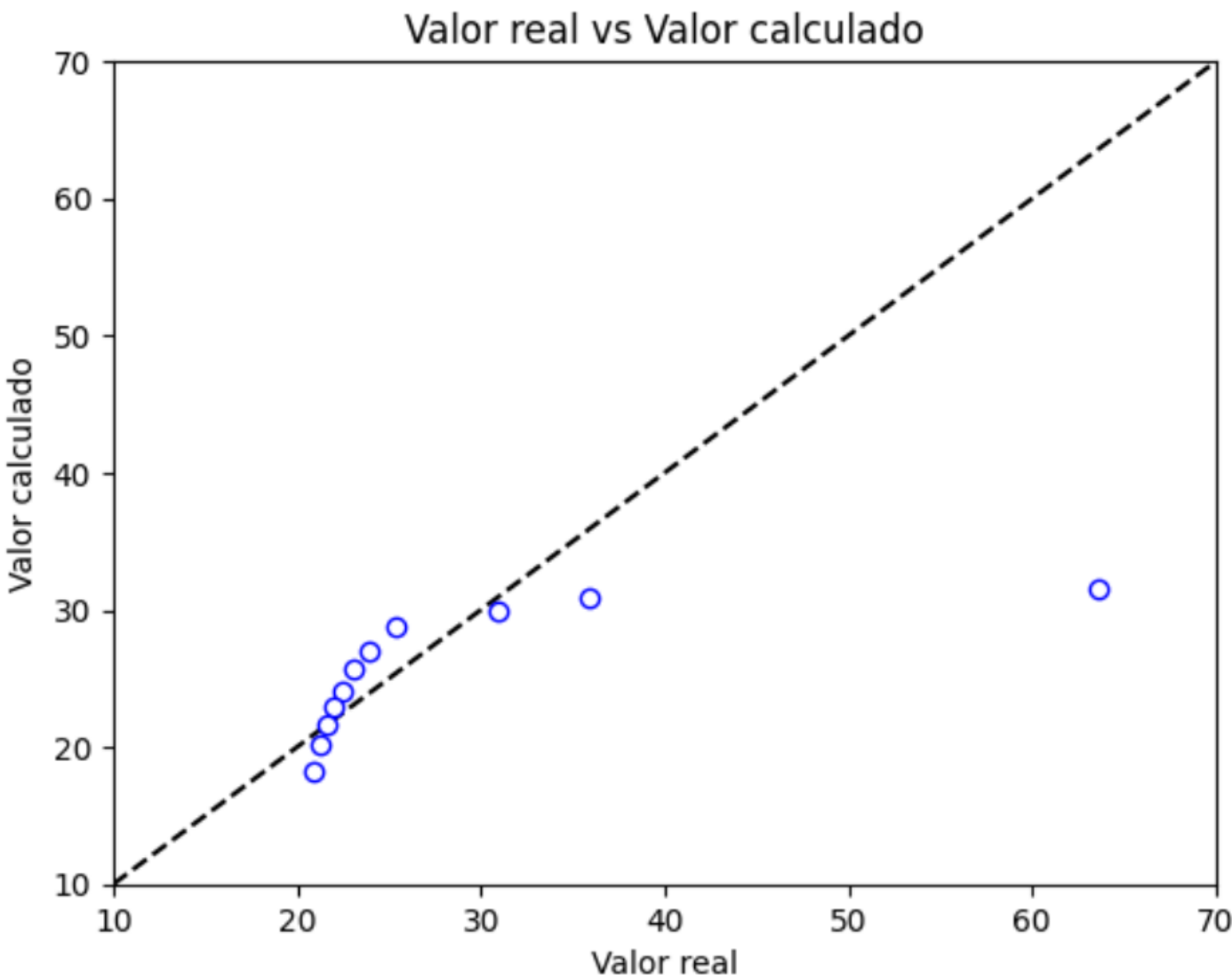
Objetivos:

- 1. Realizar una regresión lineal univariable de los datos de la muestra de “petróleo 1”.

Coeficiente “a”	-0.1453
Coeficiente “b”	32.1510
Coeficiente “c”	NA
Coeficiente de determinación ( $R^2$ )	0.4724

<https://colab.research.google.com/drive/1upkmsR7Ea9hJ17dtql5cc7g1nb3BHvyM?usp=sharing>

$$\eta = a(T) + b$$



1\_Ejemplo\_1\_CSV

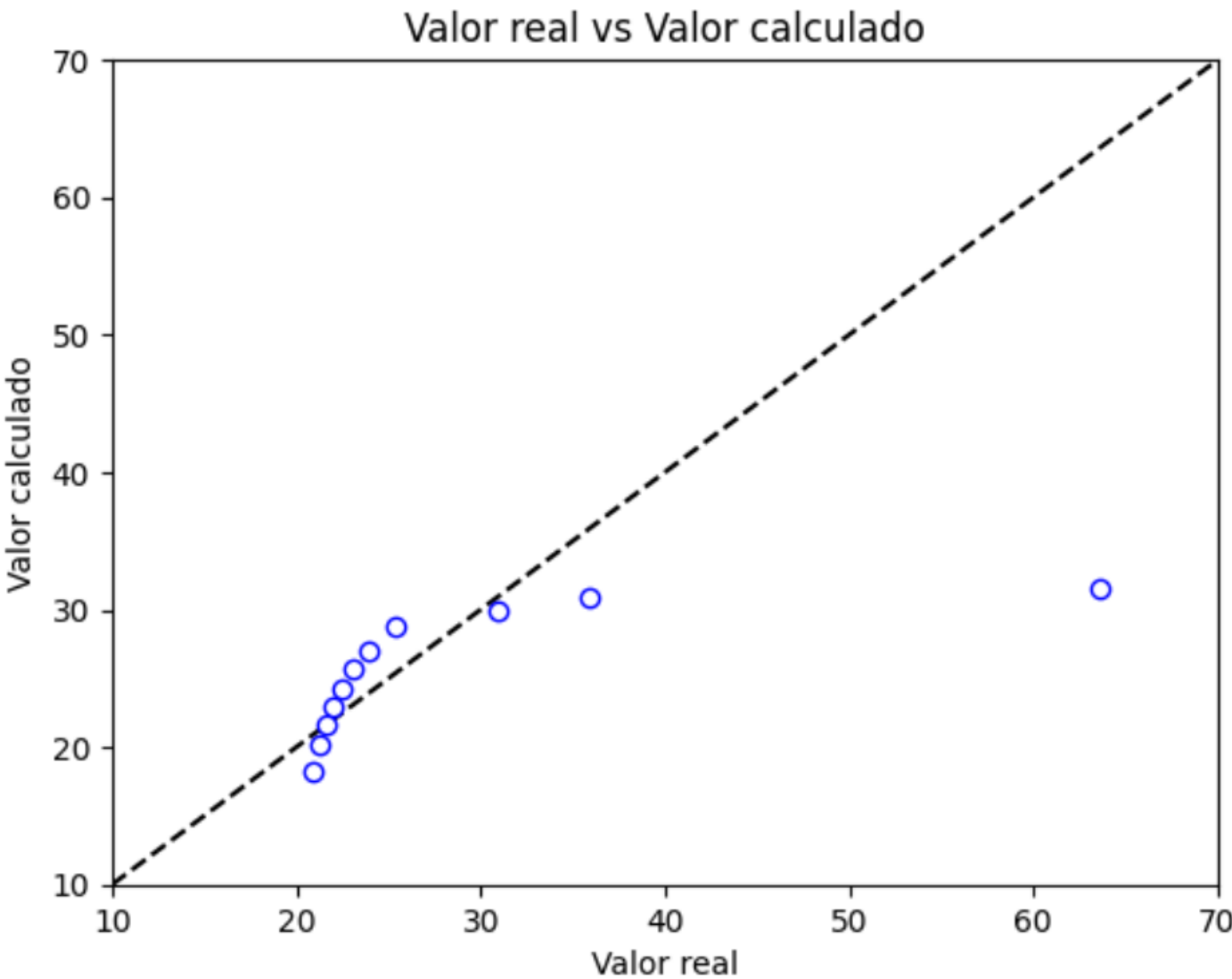
Objetivos:

- 1. Realizar una regresión lineal multivariable de los datos de la muestra de “petróleo 1”.

Coeficiente “a”	-0.1340
Coeficiente “b”	16.0561
Coeficiente “c”	16.9119
Coeficiente de determinación ( $R^2$ )	0.4730

<https://colab.research.google.com/drive/1upkmsR7Ea9hJ17dtql5cc7g1nb3BHvyM?usp=sharing>

$$\eta = a(T) + b(\rho) + c$$



1\_Ejemplo\_1\_CSV

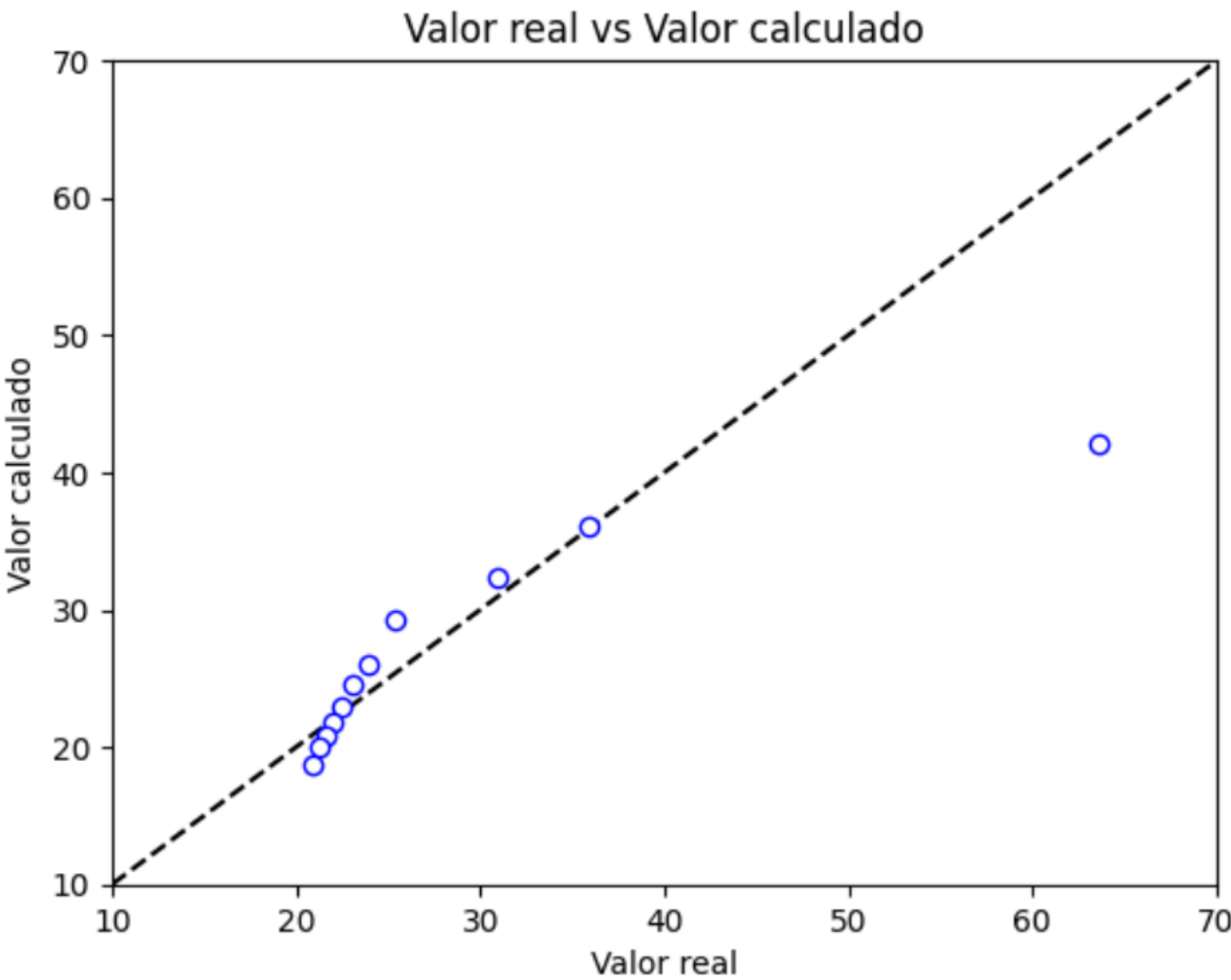
Objetivos:

- 1. Realizar una regresión logarítmica univariable de los datos de la muestra de “petróleo 1”.

Coeficiente “a”	-7.3089
Coeficiente “b”	52.1601
Coeficiente “c”	NA
Coeficiente de determinación (R²)	0.7969

<https://colab.research.google.com/drive/1upkmsR7Ea9hJ17dtql5cc7g1nb3BHvyM?usp=sharing>

$\eta = a * \ln(T) + b$



1\_Ejemplo\_1\_CSV

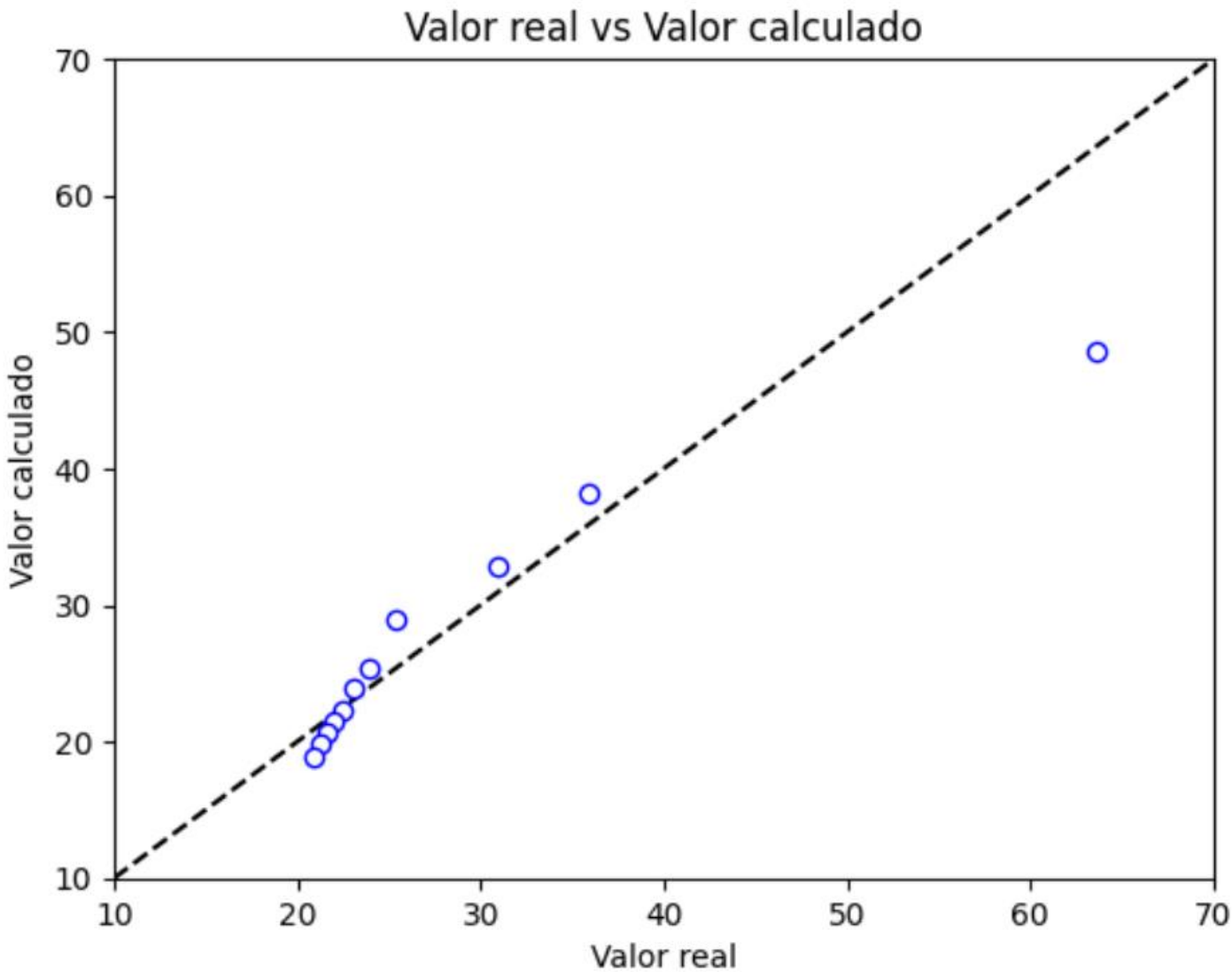
Objetivos:

- 1. Realizar una regresión potencial univariable de los datos de la muestra de “petróleo 1”.

Coeficiente “a”	73.1996
Coeficiente “b”	-0.2959
Coeficiente “c”	NA
Coeficiente de determinación ( $R^2$ )	0.8884

<https://colab.research.google.com/drive/1upkmsR7Ea9hJ17dtql5cc7g1nb3BHvyM?usp=sharing>

$\eta = a(T^b)$





1\_Ejemplo\_1\_CSV

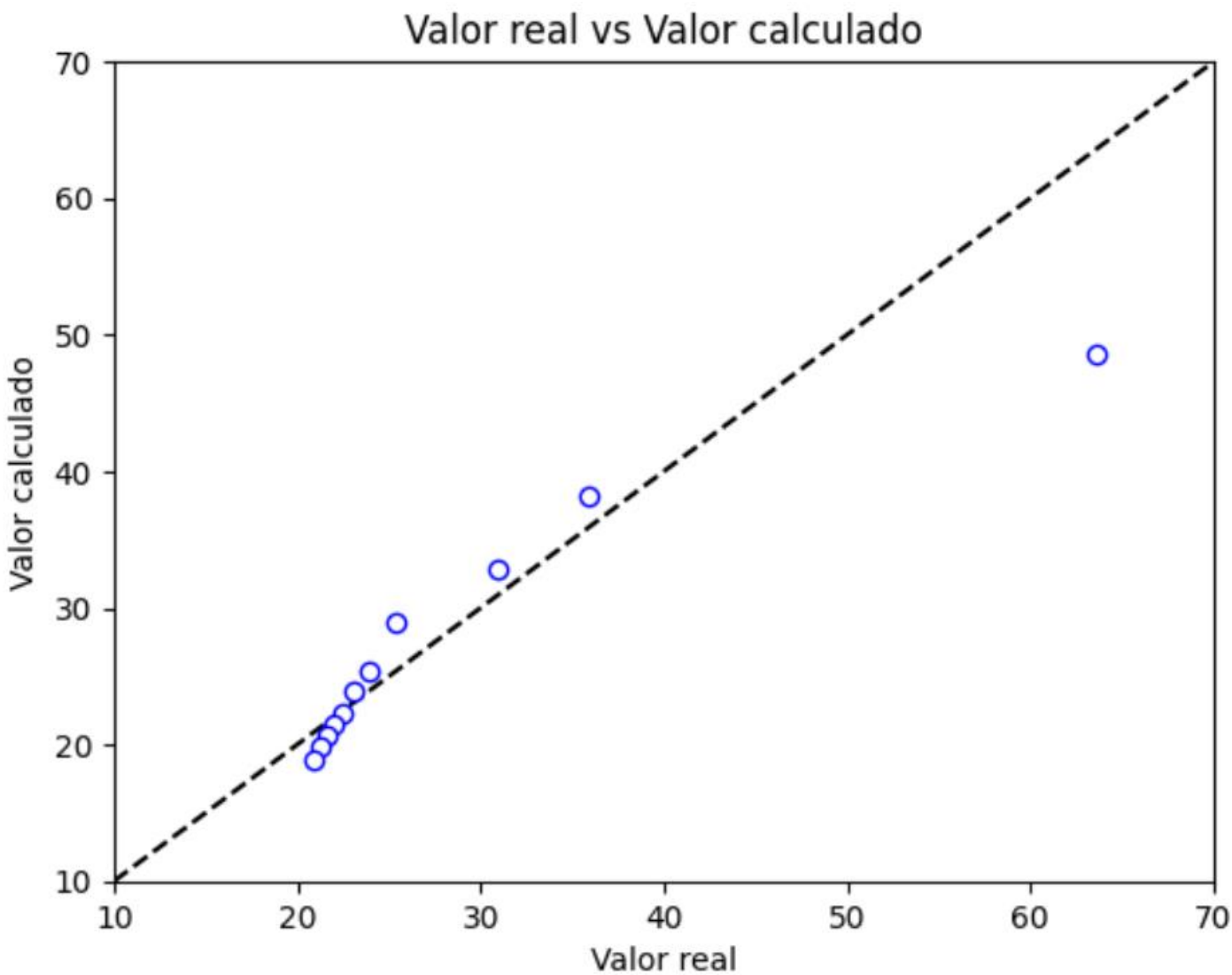
Objetivos:

- 1. Realizar una regresión potencial multivariable de los datos de la muestra de “petróleo 1”.

Coeficiente “a”	69.9724
Coeficiente “b”	-0.5322
Coeficiente “c”	-10.140
Coeficiente de determinación (R²)	0.9755

<https://colab.research.google.com/drive/1upkmsR7Ea9hJ17dtql5cc7g1nb3BHvyM?usp=sharing>

$\eta = a(T^b)(\rho^c)$



2\_Ejemplo\_2\_CSV

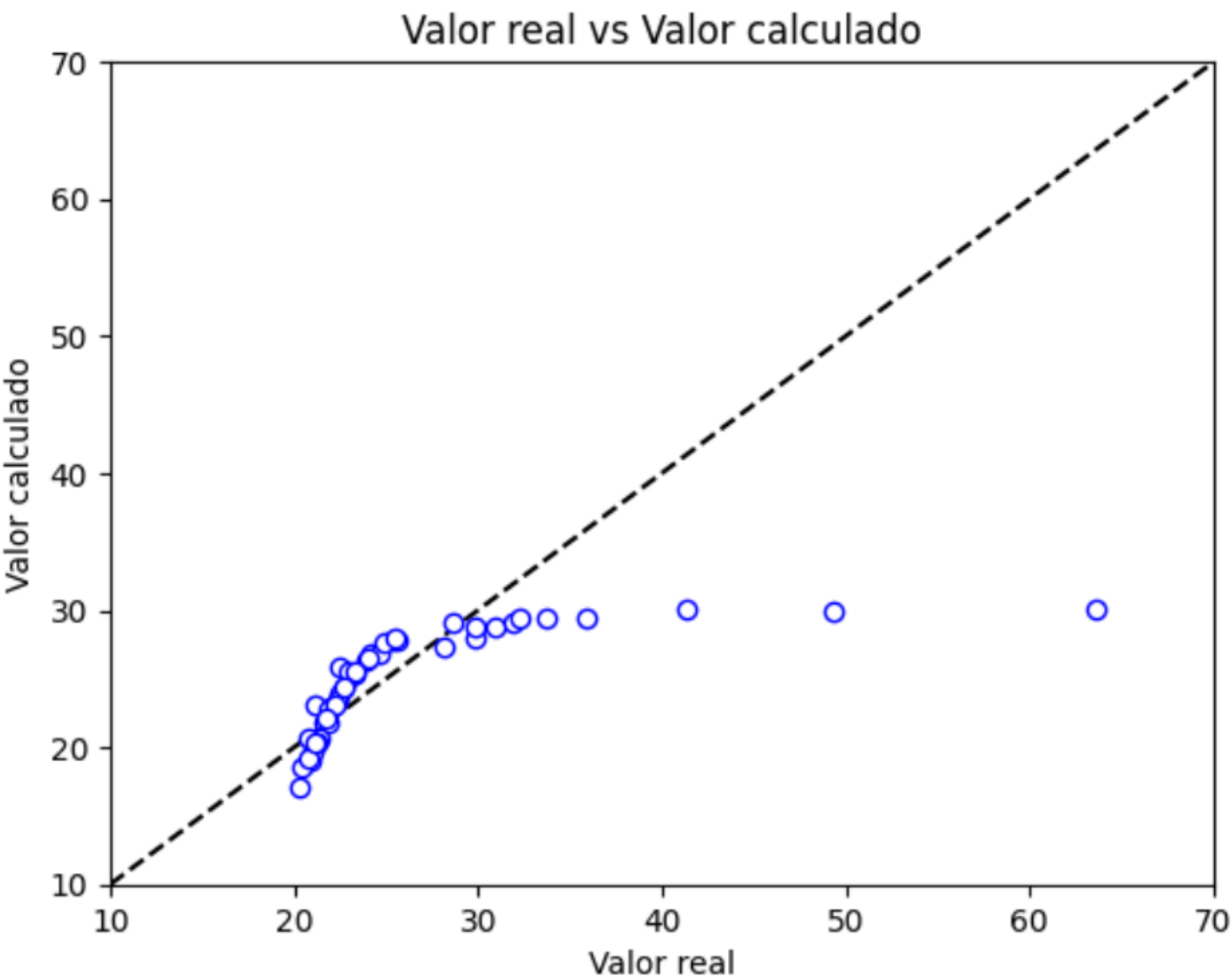
Objetivos:

- 1. Realizar una regresión lineal multivariable de los datos de las muestras de “petróleo 1”, “petróleo 2”, “petróleo 10”, “petróleo 15” y “petróleo 27”.

Coeficiente “a”	-0.07514
Coeficiente “b”	64.7599
Coeficiente “c”	-30.8900
Coeficiente de determinación ( $R^2$ )	0.2975

<https://colab.research.google.com/drive/14emMFnknI8UsH2eUmnsV5EEoxjhwXGA4?usp=sharing>

$\eta = a(T) + b(\rho) + c$



2\_Ejemplo\_2\_CSV

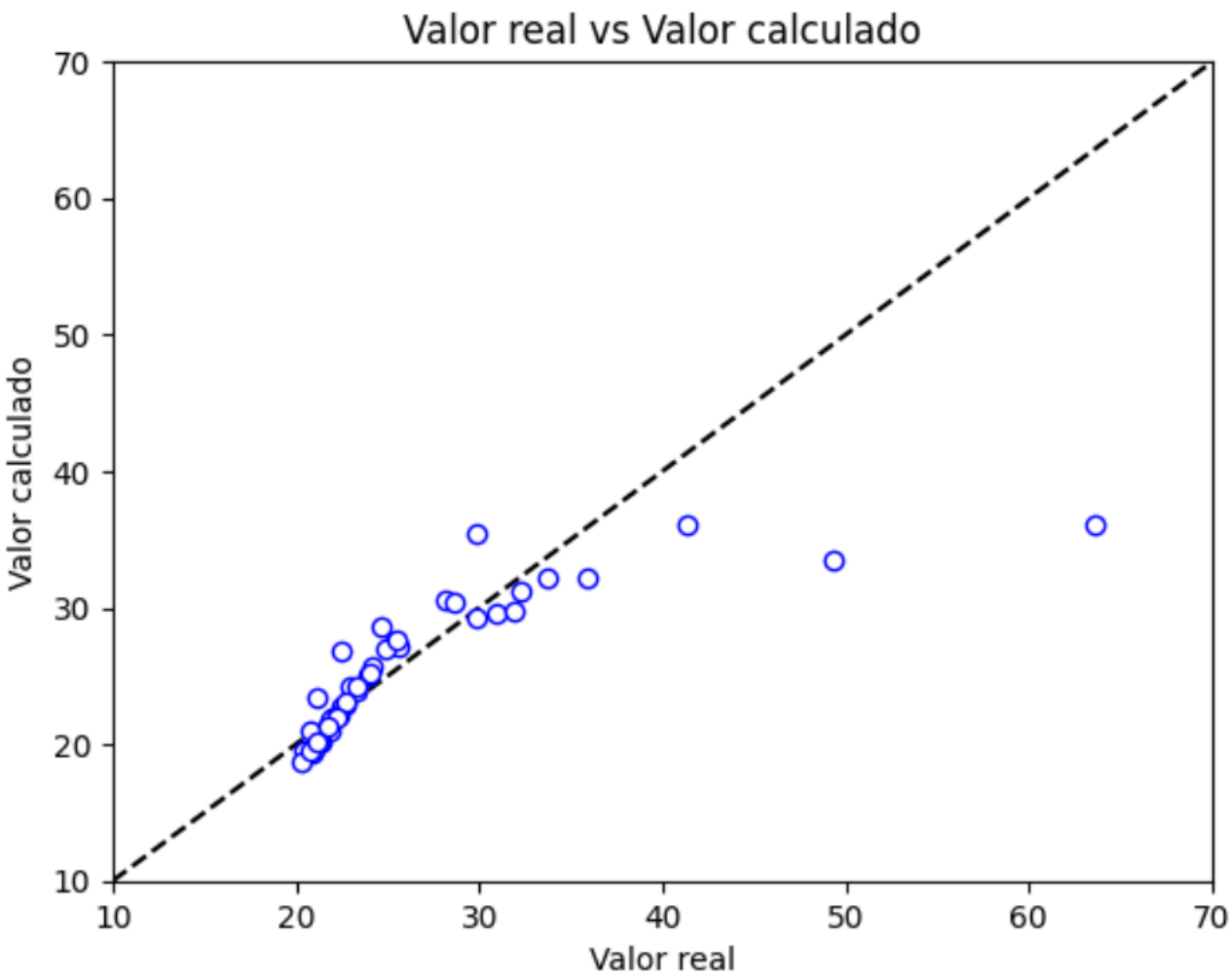
Objetivos:

- 1. Realizar una regresión **logarítmica multivariable** de los datos de las muestras de “petróleo 1”, “petróleo 2”, “petróleo 10”, “petróleo 15” y “petróleo 27”.

Coeficiente “a”	-4.8021
Coeficiente “b”	21.0702
Coeficiente “c”	43.9547
Coeficiente de determinación ( $R^2$ )	0.4674

<https://colab.research.google.com/drive/14emMFnknI8UsH2eUmnsv5EEoxjhwXGA4?usp=sharing>

$$\eta = a * \ln(T) + b * \ln(\rho) + c$$



2\_Ejemplo\_2\_CSV

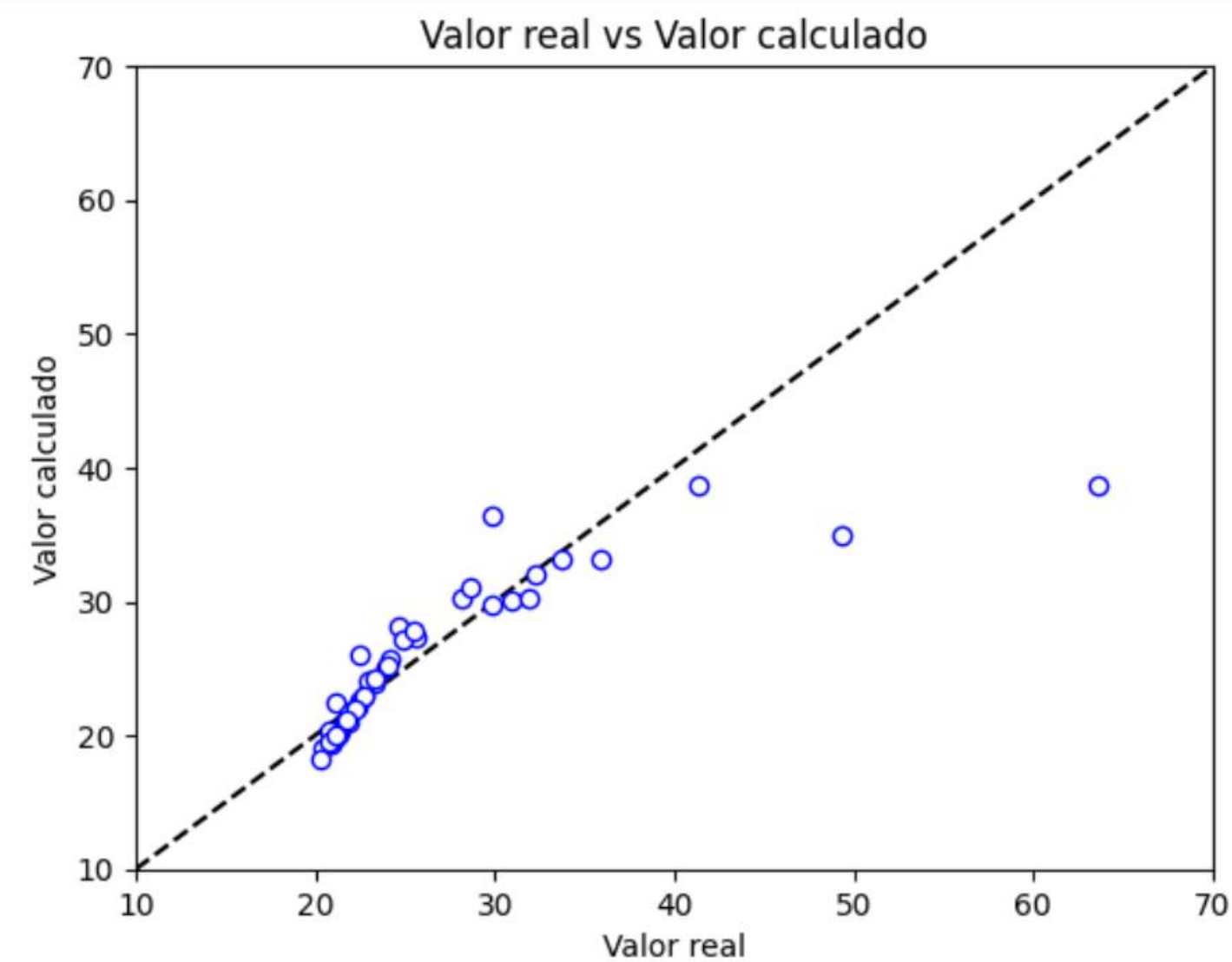
Objetivos:

- 1. Realizar una regresión **potencial multivariable** de los datos de las muestras de “petróleo 1”, “petróleo 2”, “petróleo 10”, “petróleo 15” y “petróleo 27”.

Coeficiente “a”	54.5642
Coeficiente “b”	-0.1797
Coeficiente “c”	1.7244
Coeficiente de determinación ( $R^2$ )	0.5199

<https://colab.research.google.com/drive/14emMFnknI8UsH2eUmnsv5EEoxjhwXGA4?usp=sharing>

$\eta = a(T^b)(\rho^c)$



2\_Ejemplo\_2\_CSV

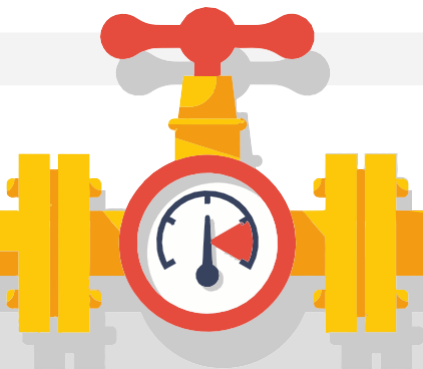
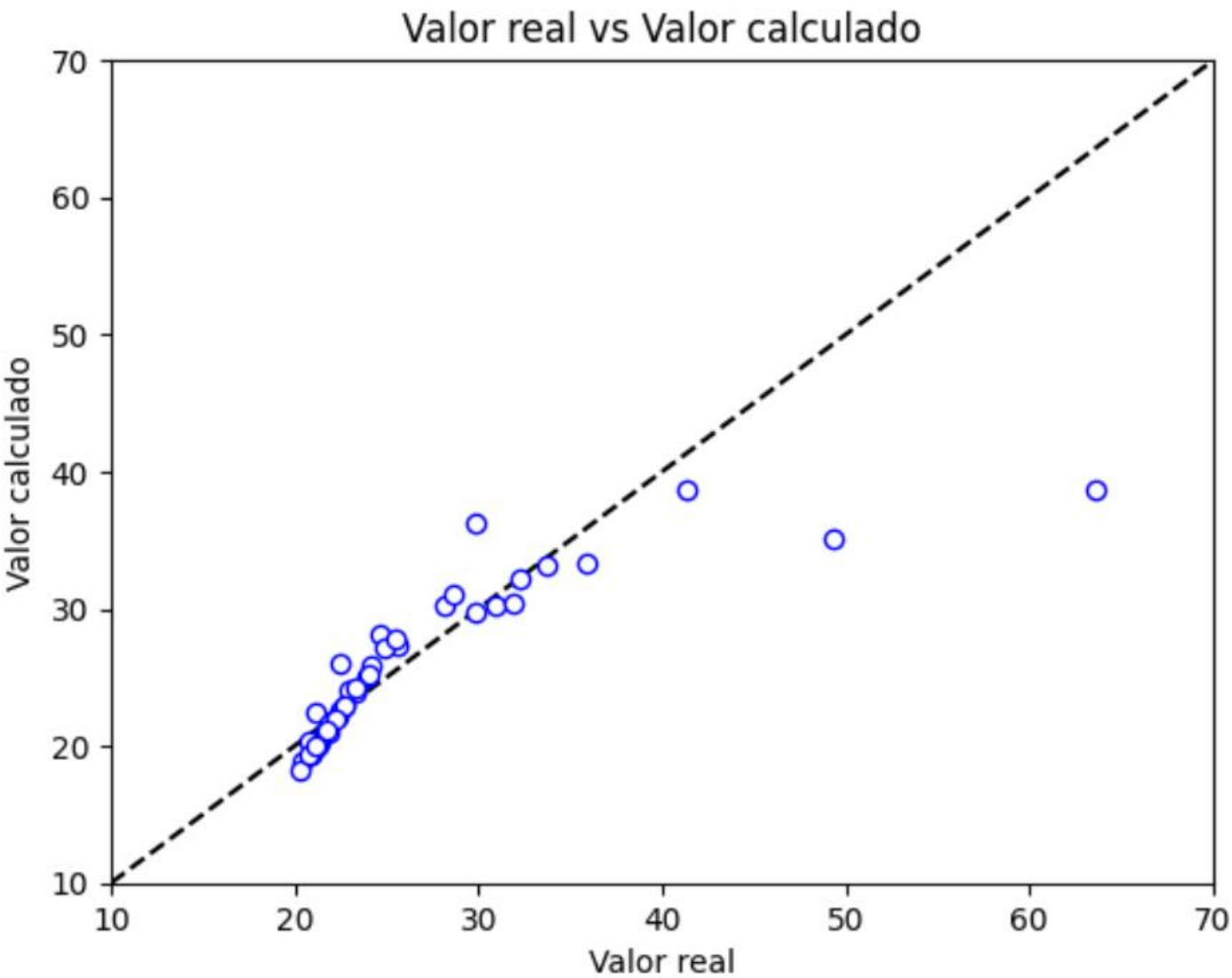
Objetivos:

- 1. Realizar una regresión tipo Arrhenius de los datos de las muestras de “petróleo 1”, “petróleo 2”, “petróleo 10”, “petróleo 15” y “petróleo 27”.

Coeficiente “a”	7.6668
Coeficiente “b”	-0.1784
Coeficiente “c”	1.9718
Coeficiente de determinación (R²)	0.5211

<https://colab.research.google.com/drive/14emMFnknI8UsH2eUmnsv5EEoxjhwXGA4?usp=sharing>

$\eta = a * e^{(b(T) + c(\rho))}$



2\_Ejemplo\_2\_CSV

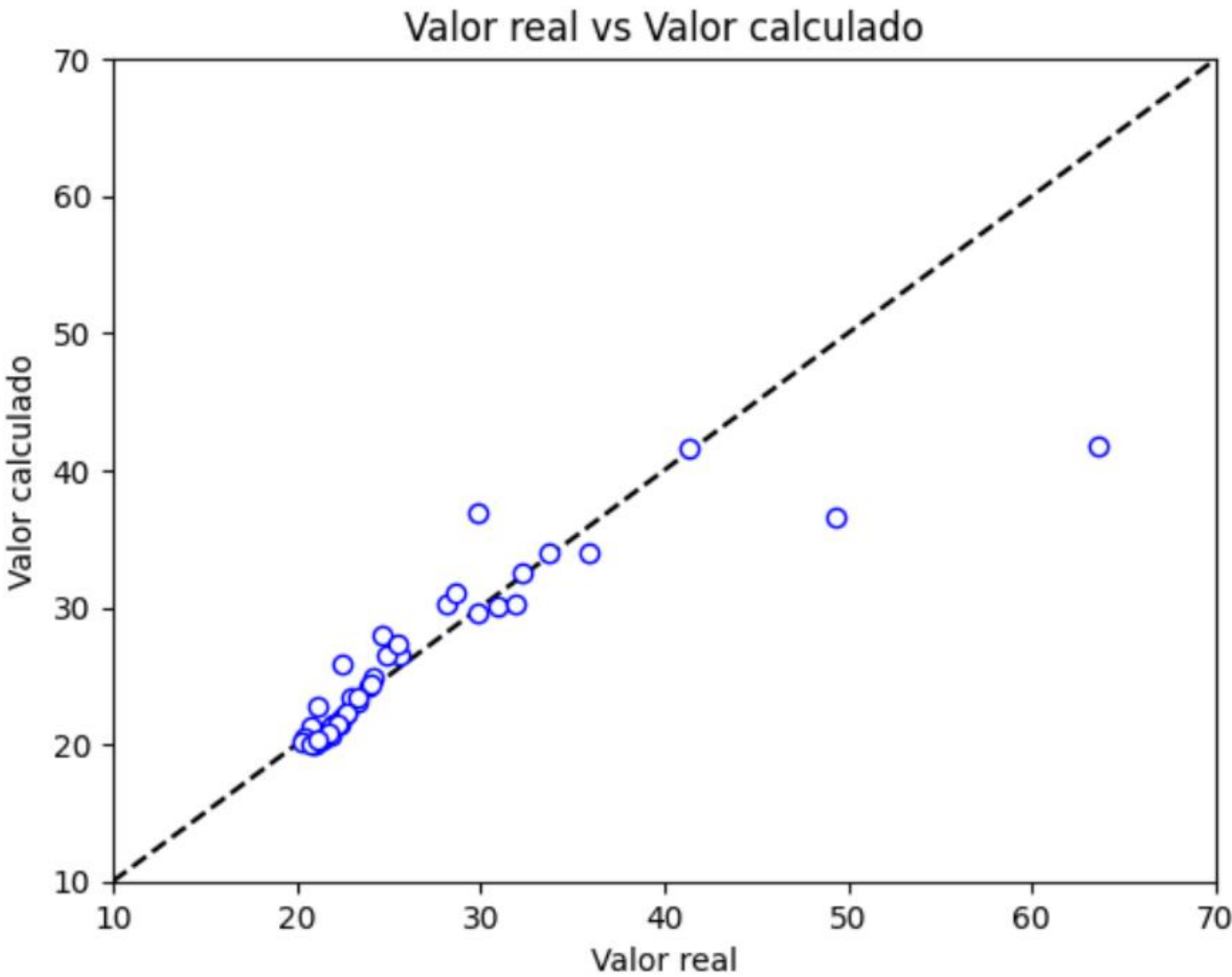
Objetivos:

- 1. Realizar una regresión tipo Arrhenius\* de los datos de las muestras de “petróleo 1”, “petróleo 2”, “petróleo 10”, “petróleo 15” y “petróleo 27”.

Coeficiente “a”	0.3963
Coeficiente “b”	0.9246
Coeficiente “c”	5.2675
Coeficiente “d”	-1.2287
Coeficiente de determinación (R²)	0.5932

<https://colab.research.google.com/drive/14emMFnknI8UsH2eUmnsv5EEoxjhwXGA4?usp=sharing>

$$\eta = a * e^{(b(T) + c(\rho) + d(T)(\rho))}$$





3\_Ejemplo\_3\_CSV

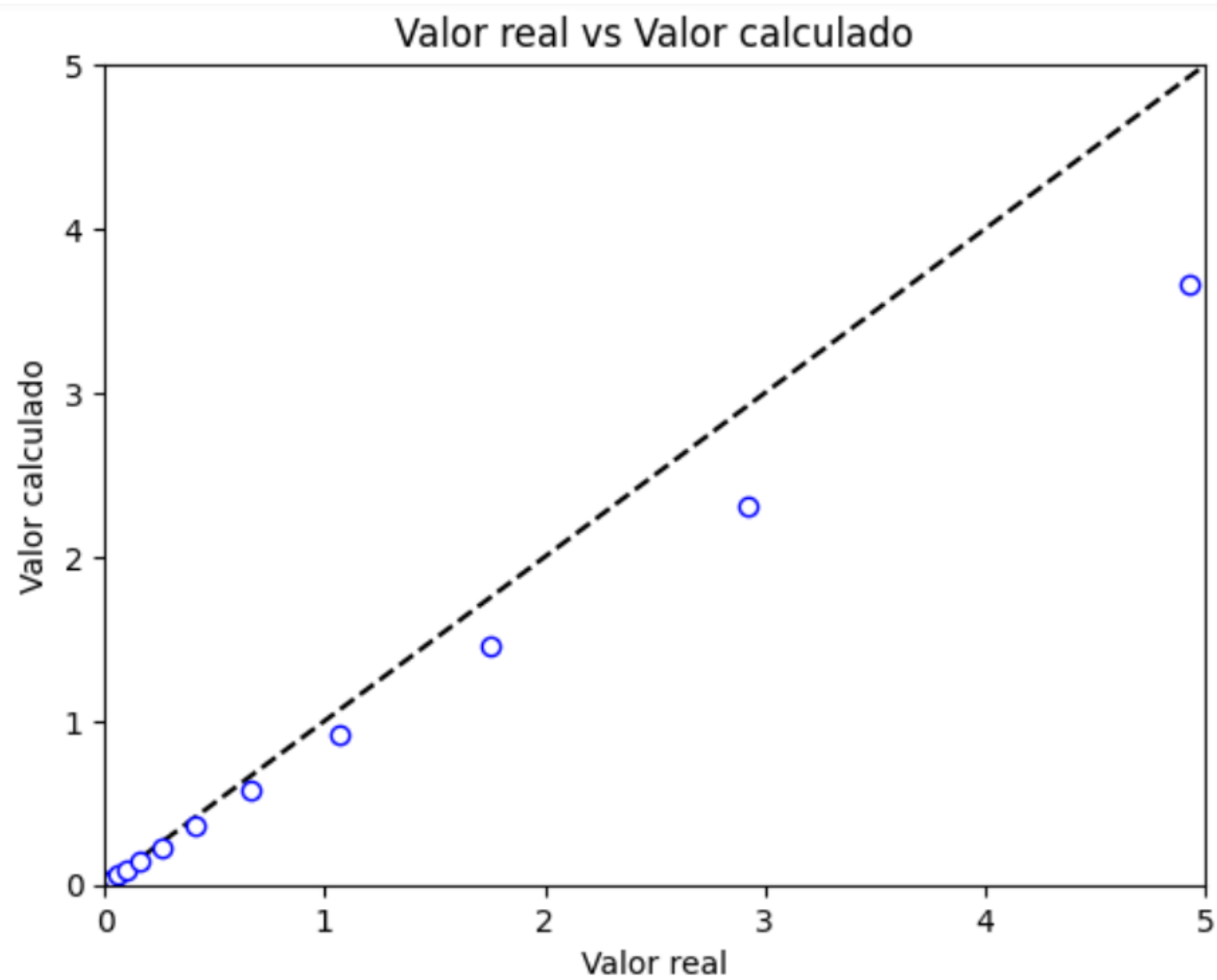
Objetivos:

- 1. Realizar una regresión lineal multivariable de los datos de la muestra de “petróleo 3” del pozo #105 a temperatura de 25 °C.

Coeficiente “a”	0.0036
Coeficiente “b”	-0.3729
Coeficiente “c”	4.8325
Coeficiente “d”	4.8375
Coeficiente de determinación ( $R^2$ )	0.9975

<https://colab.research.google.com/drive/12UdzwXV5on3jcLaquclmavt4omfZVWBO?usp=sharing>

$$\sigma = a * (\dot{\gamma}) + b(T) + c(\rho) + d$$





3\_Ejemplo\_3\_CSV

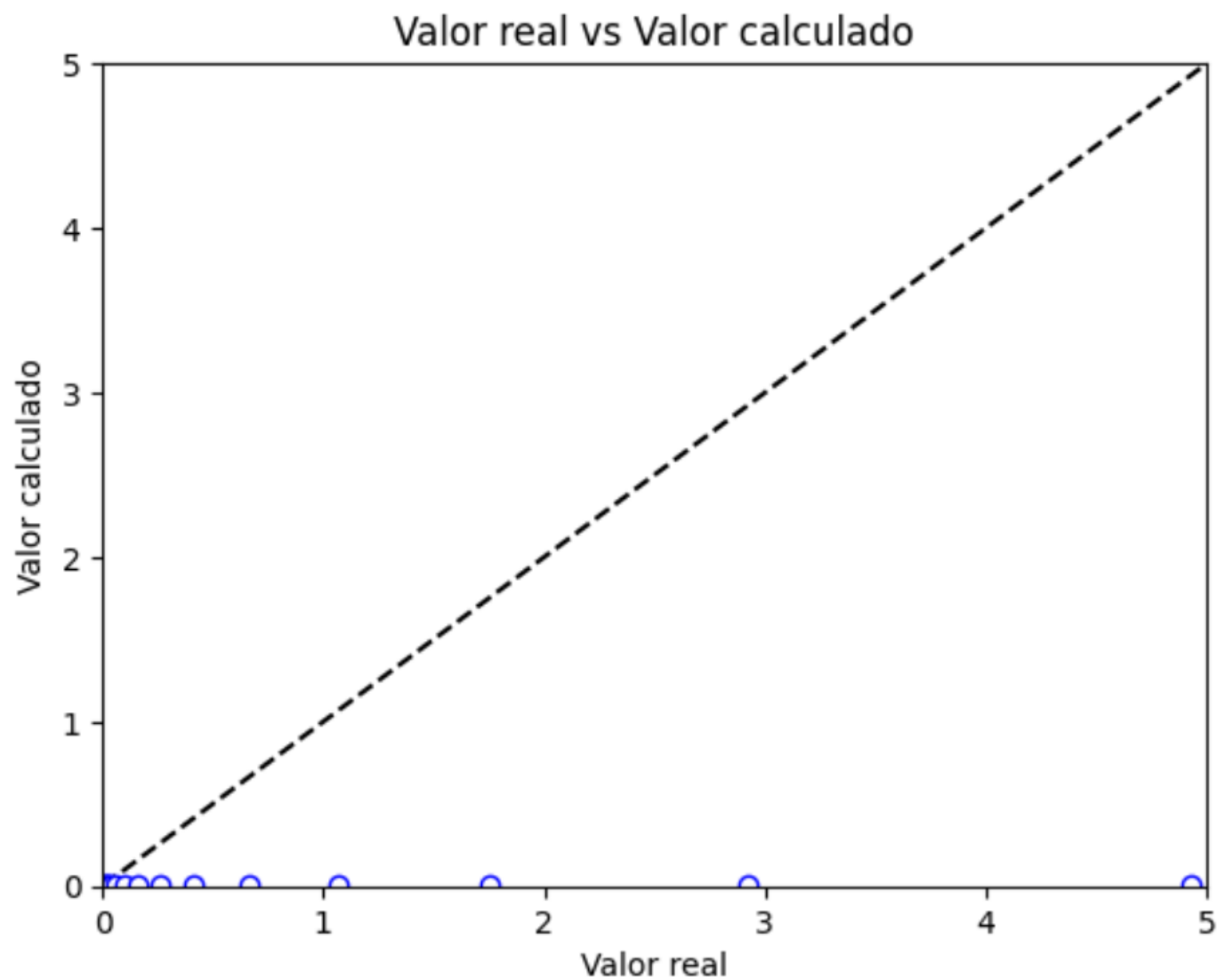
Objetivos:

- 1. Realizar una regresión **logarítmica multivariable** de los datos de la muestra de “petróleo 3” del pozo #105 a temperatura de 25° C.

Coeficiente “a”	0.0013
Coeficiente “b”	0.0007
Coeficiente “c”	6.0714
Coeficiente “d”	0.0002
Coeficiente de determinación ( $R^2$ )	0.4945

<https://colab.research.google.com/drive/12UdzwXV5on3jcLaquclmavt4omfZVWBO?usp=sharing>

$$\sigma = a * \ln(\gamma') + b * \ln(T) + c * \ln(\rho) + d$$



3\_Ejemplo\_3\_CSV

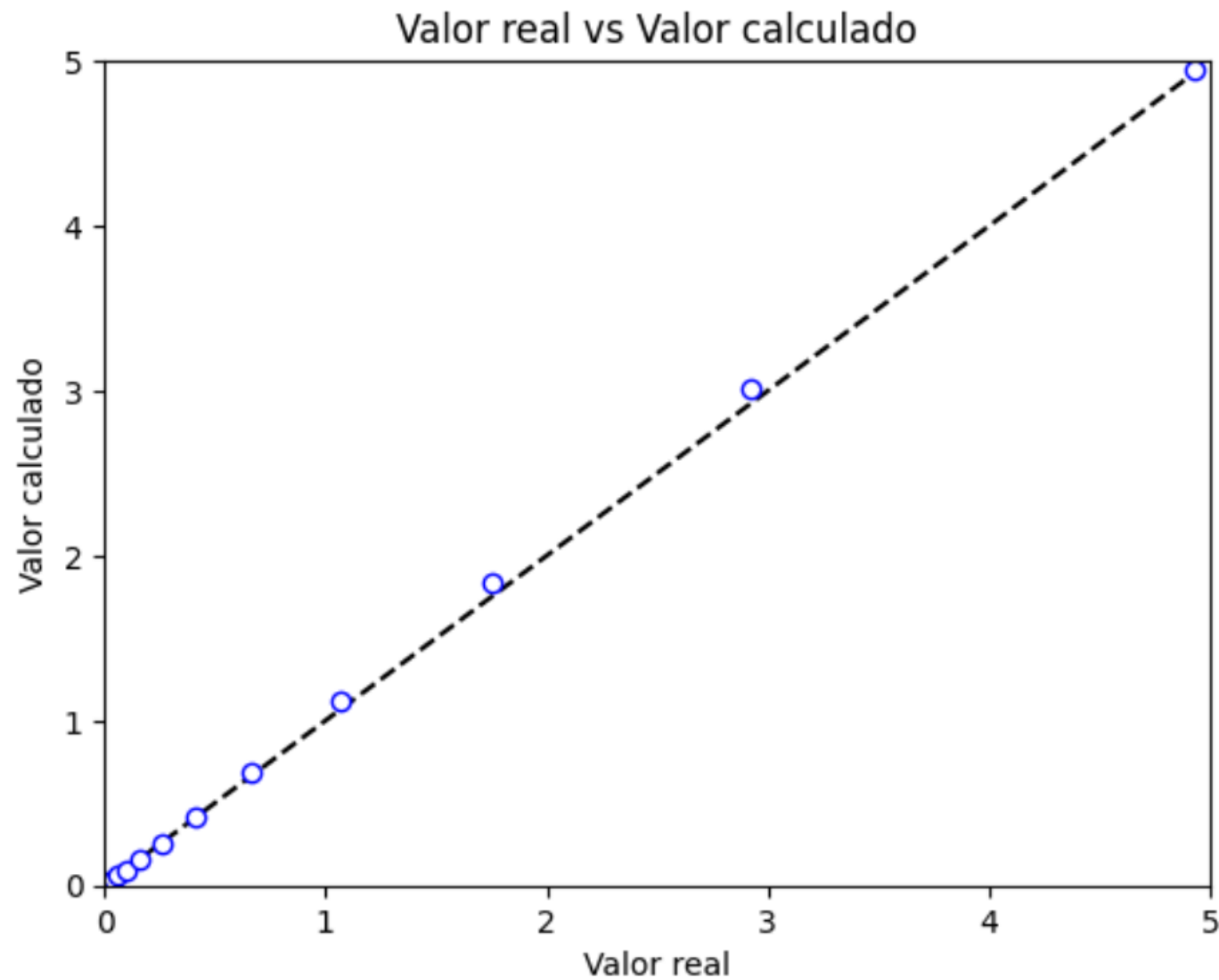
Objetivos:

- 1. Realizar una regresión **potencial multivariable** de los datos de la muestra de “petróleo 3” del pozo #105 a temperatura de 25° C.

Coeficiente “a”	0.0053
Coeficiente “b”	1.077
Coeficiente “c”	-0.190
Coeficiente “d”	0.0022
Coeficiente de determinación ( $R^2$ )	0.9996

<https://colab.research.google.com/drive/12UdzwXV5on3jcLaquclmavt4omfZVWBO?usp=sharing>

$$\sigma = a * (\dot{\gamma})^b (T)^c (\rho)^d$$



4\_Ejemplo\_4\_CSV

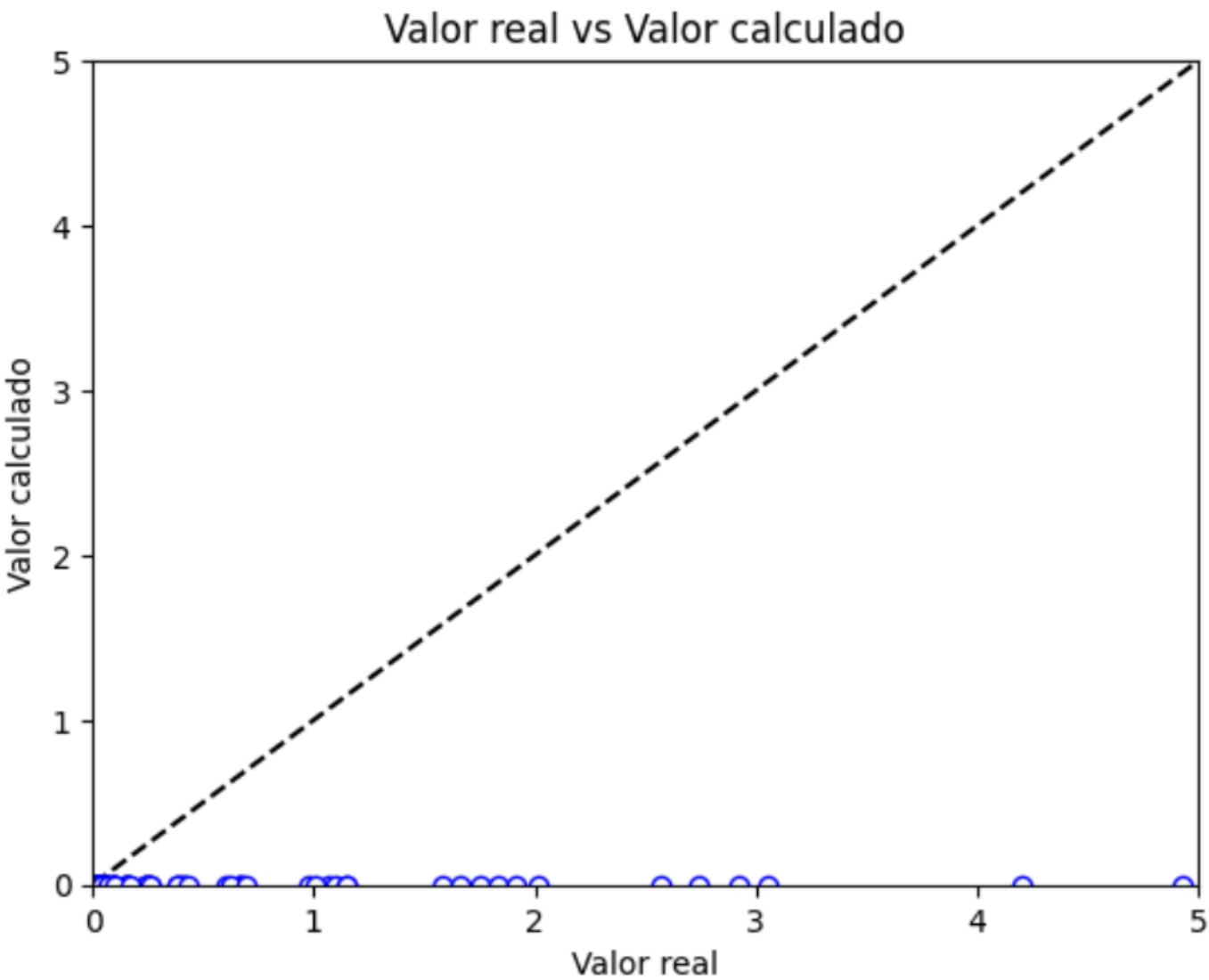
Objetivos:

- 1. Realizar una regresión lineal multivariable de los datos de todas las muestras del “Campo 1”.

Coeficiente “a”	0
Coeficiente “b”	0
Coeficiente “c”	0
Coeficiente “d”	0
Coeficiente de determinación ( $R^2$ )	0

<https://colab.research.google.com/drive/12UdzwXV5on3jcLaquclmavt4omfZVWB0?usp=sharing>

$$\eta = a(\dot{\gamma}) + b(T) + c(\rho) + d$$



4\_Ejemplo\_4\_CSV

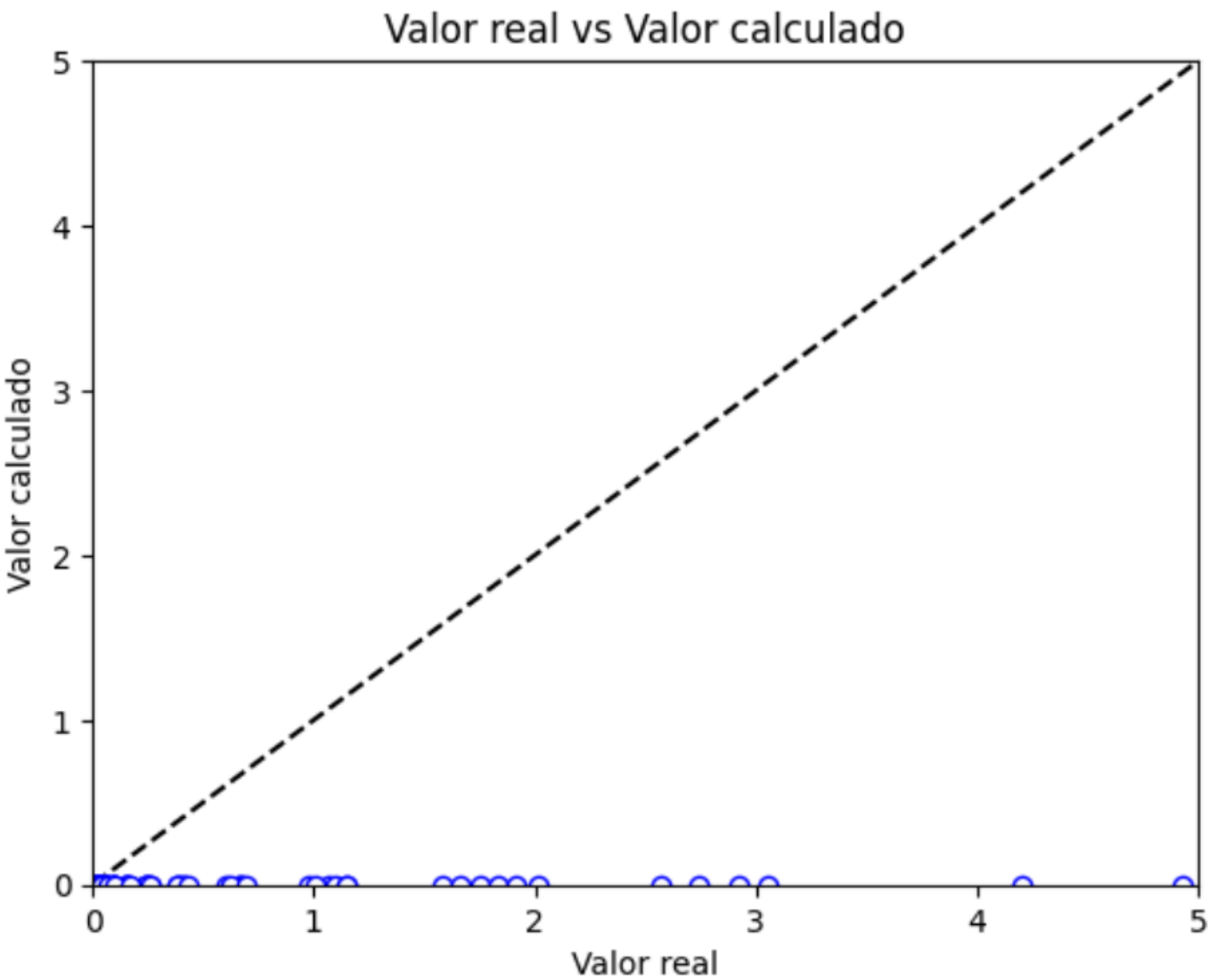
Objetivos:

- 1. Realizar una regresión **logarítmica multivariable** de los datos de todas las muestras del “Campo 1”.

Coeficiente “a”	0.0013
Coeficiente “b”	0.0009
Coeficiente “c”	0.0010
Coeficiente “d”	-0.0011
Coeficiente de determinación ( $R^2$ )	0.4002

<https://colab.research.google.com/drive/12UdzwXV5on3jcLaquclmavt4omfZVWB0?usp=sharing>

$$\eta = a(\dot{\gamma}) + b(T) + c(\rho) + d$$



4\_Ejemplo\_4\_CSV

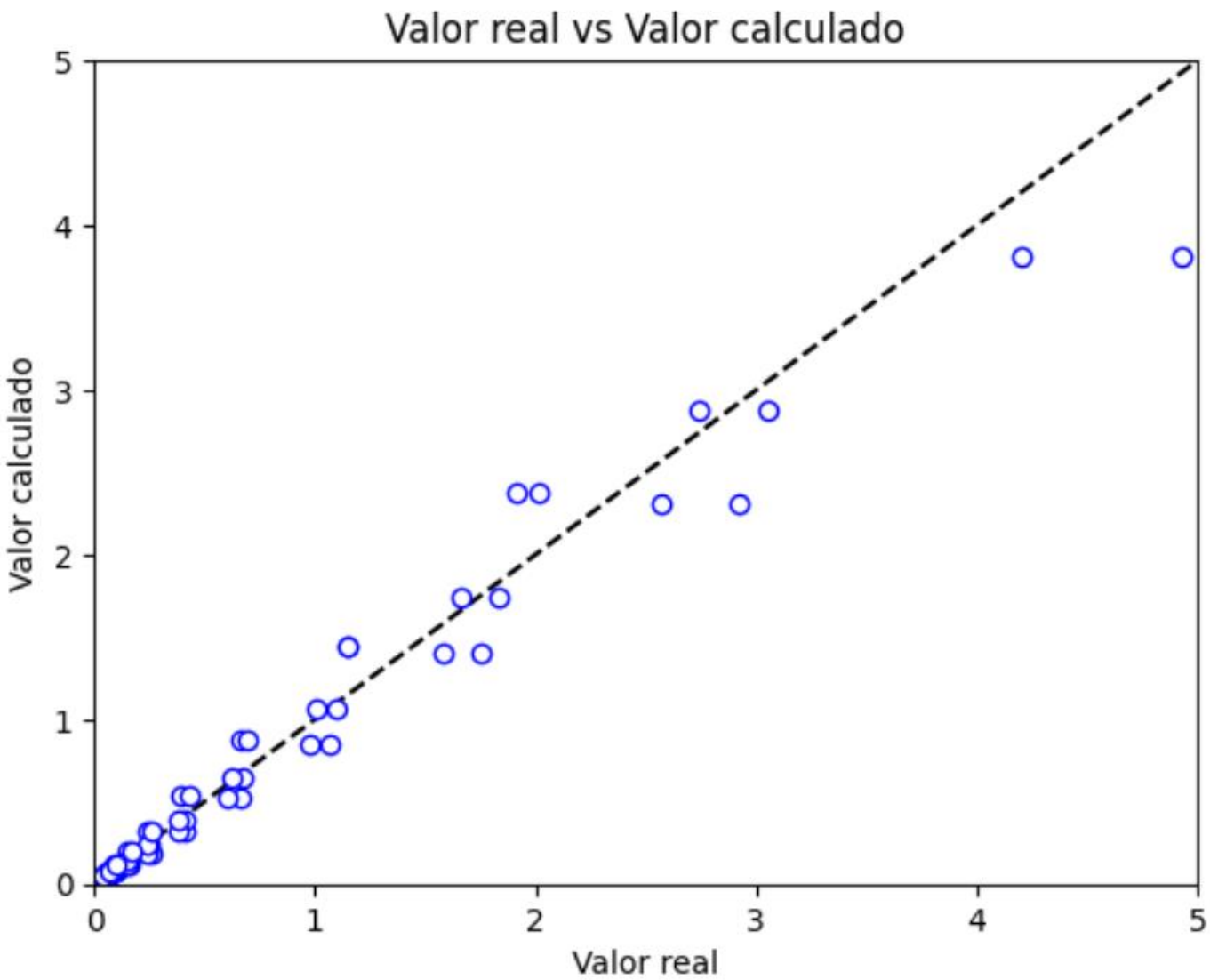
Objetivos:

- 1. Realizar una regresión **potencial multivariable** de los datos de todas las muestras del “Campo 1”.

Coeficiente “a”	0.0040
Coeficiente “b”	1.0830
Coeficiente “c”	-0.4035
Coeficiente “d”	-8.91420
Coeficiente de determinación ( $R^2$ )	0.971

<https://colab.research.google.com/drive/12UdzwXV5on3jcLaquclmavt4omfZVWBO?usp=sharing>

$$\sigma = a * (\dot{\gamma})^b (T)^c (\rho)^d$$







Universidad Nacional de Colombia  
SPE Student Chapter

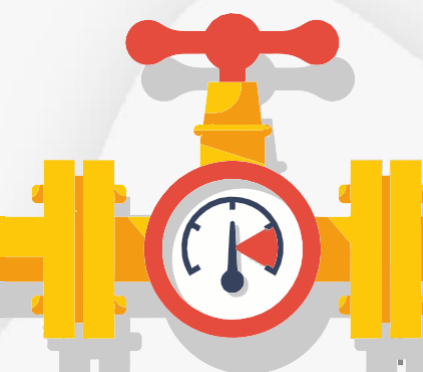
PETROIntelligence.com

## SESIÓN 4 REGRESIONES APLICADAS A LA INGENIERÍA DE PETRÓLEOS (CON PYTHON)

Misael Edgar Zepeda Díaz

[mzepeda@petrointelligence.com](mailto:mzepeda@petrointelligence.com)  
(+52 55 1350 7432)

Agosto 2023

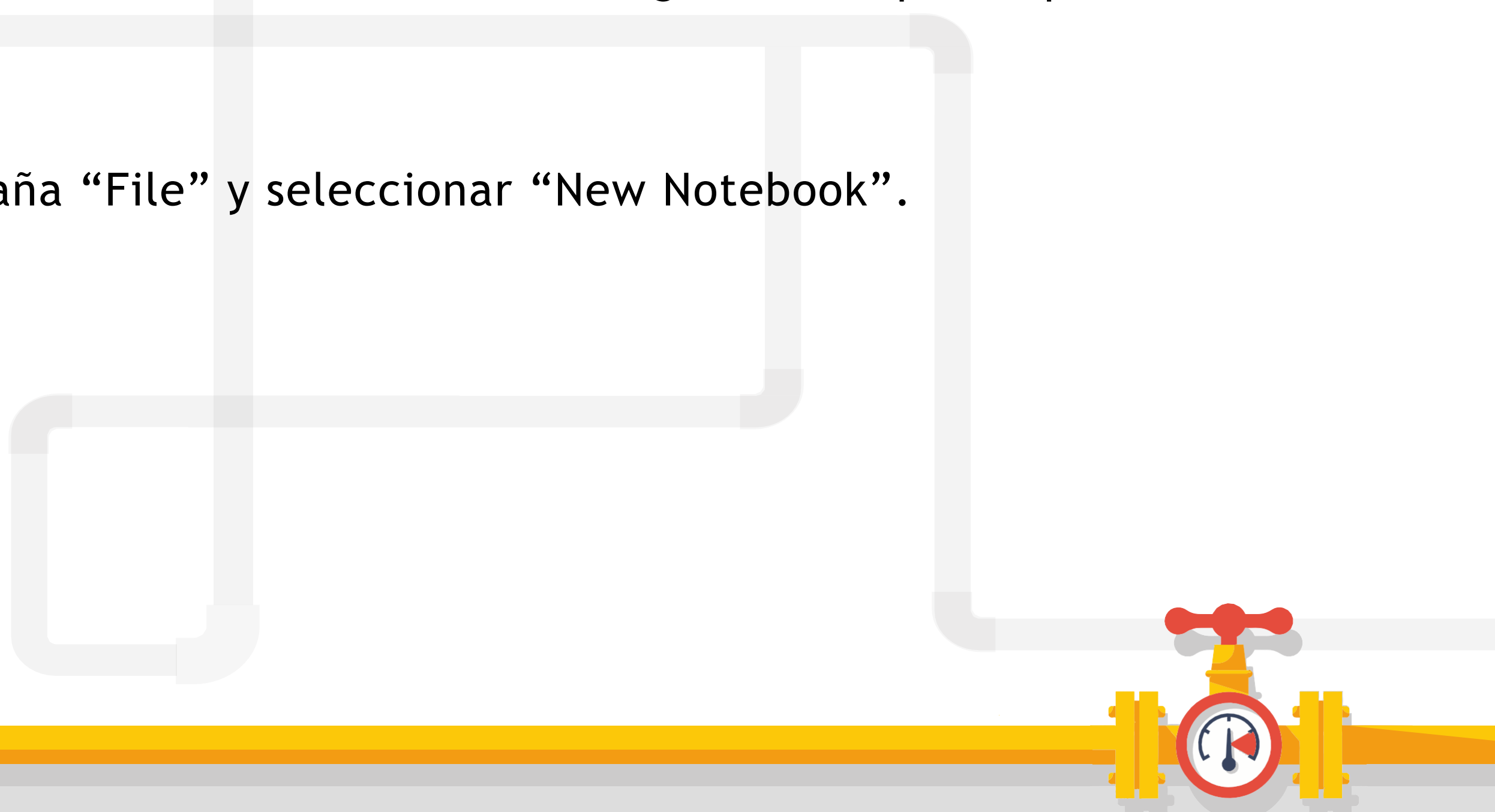


Título	Horas
1. Introducción -----	1
2. Manejo de datos -----	1
3. Python -----	2
4. Regresiones -----	1
5. Análisis de los resultados -----	1
Total =	6



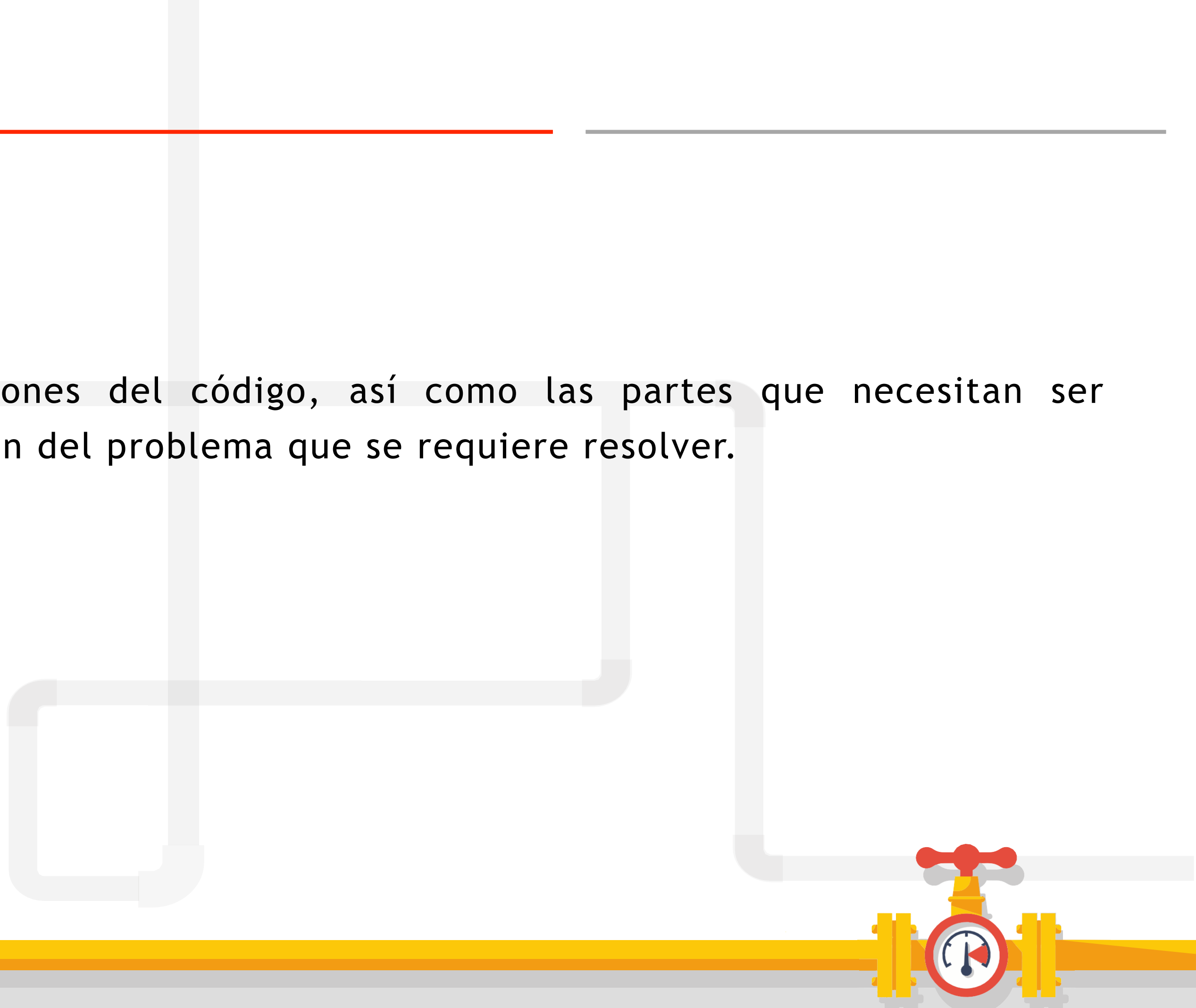
Se recomienda que se copien los códigos a un nuevo *notebook* en Google Colab que le pertenezca al usuario.

El procedimiento es dando *click* en la pestaña “File” y seleccionar “New Notebook”.



### Estructura del código

A continuación, se muestran las secciones del código, así como las partes que necesitan ser modificadas de acuerdo a la configuración del problema que se requiere resolver.



## 1. IMPORTACIÓN DE LIBRERÍAS

Las librerías son un conjunto de archivos (códigos) que se utilizan para facilitar el desarrollo de software.

Éstas proporcionan funcionalidades y módulos que facilitan la solución de problemas estándar.

Esta sección no requiere modificaciones.



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

## 2. CARGA DE DATOS DE ENTRADA

Una vez que la base de datos de entrada está preparada, se procede a la carga de datos en el programa.



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

Se requiere archivo en formato “CSV”.

Esta sección requiere modificaciones:

- Colocar la ubicación del archivo de entrada.

### 3. ASIGNACIÓN DE LAS COLUMNAS DEL ARCHIVO CARGADO EN ARREGLOS DEL PROGRAMA

Esta sección requiere modificaciones:

- Colocar los nombres de las cabeceras de las columnas del archivo de entrada en cada uno de los arreglos definidos en el programa.

	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



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

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

## 4. INGRESAR EL MODELO

Se declara el nombre y cantidad de coeficientes que se utilizarán. Asimismo, se declara el modelo.

Esta sección requiere modificaciones:

- Se escribe el modelo al que se requiere aproximar la regresión.

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



```
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
```



## 5. DEFINICIÓN DE FUNCIÓN A OPTIMIZAR Y VALORES INICIALES DE LOS COEFICIENTES

Se realiza la definición de la función a optimizar, así como los valores iniciales de los coeficientes de la regresión.

Esta sección no requiere modificaciones.

```
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
```



## 6. DEFINICIÓN DEL DOMINIO DE LOS COEFICIENTES DE REGRESIÓN Y MÉTODO DE SOLUCIÓN

Se realiza la definición del dominio de los coeficientes de regresión.

Se define el *solver* y la solución de la regresión.

Esta sección requiere modificación a las fronteras de los coeficientes si los resultados así lo requieren (prueba y error).

```
# 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. IMPRESIÓN DE LOS COEFICIENTES DE REGRESIÓN Y $R^2$

Se imprimen los resultados de la solución y el  $R^2$ .

Esta sección no requiere modificación.

```
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)
```

## 8. GRÁFICA DE LOS RESULTADOS

Se grafican los valores obtenidos vs los valores calculados, así como la línea referencia.

Esta sección requiere modificación.

Se modifica el dominio de la línea de referencia para poder visualizar el rango completo de los resultados.



```
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()
```

## EJEMPLO DE APLICACIÓN

1. Variable objetivo: viscosidad [cP].
2. Variables independientes: velocidad de corte [1/s], temperatura [°C] y densidad [g/cm<sup>3</sup>].
3. El set de datos de entrada es una muestra de 10 muestras diferentes de petróleo.
4. Las temperaturas de los datos son de 25, 50 y 80 °C.
5. Las velocidades de corte van desde 2.51 hasta 1,000 [1/s].
6. Las densidades van desde 0.77 a 0.92 [g/cm<sup>3</sup>].
7. Se propone un modelo exponencial de la forma:

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

EJEMPLO DE APLICACIÓN

- 8. Preparar los datos de entrada en una hoja de Excel (extraer datos desde “3\_Viscosidad\_Base.csv”)
- 9. Recuerda que las primeras columnas del archivo deben de ser las variables independientes.
- 10. Recuerda que la última columna debe corresponder a la variable dependiente.
- 11. Guarda el archivo con formato: CSV (separado por comas).

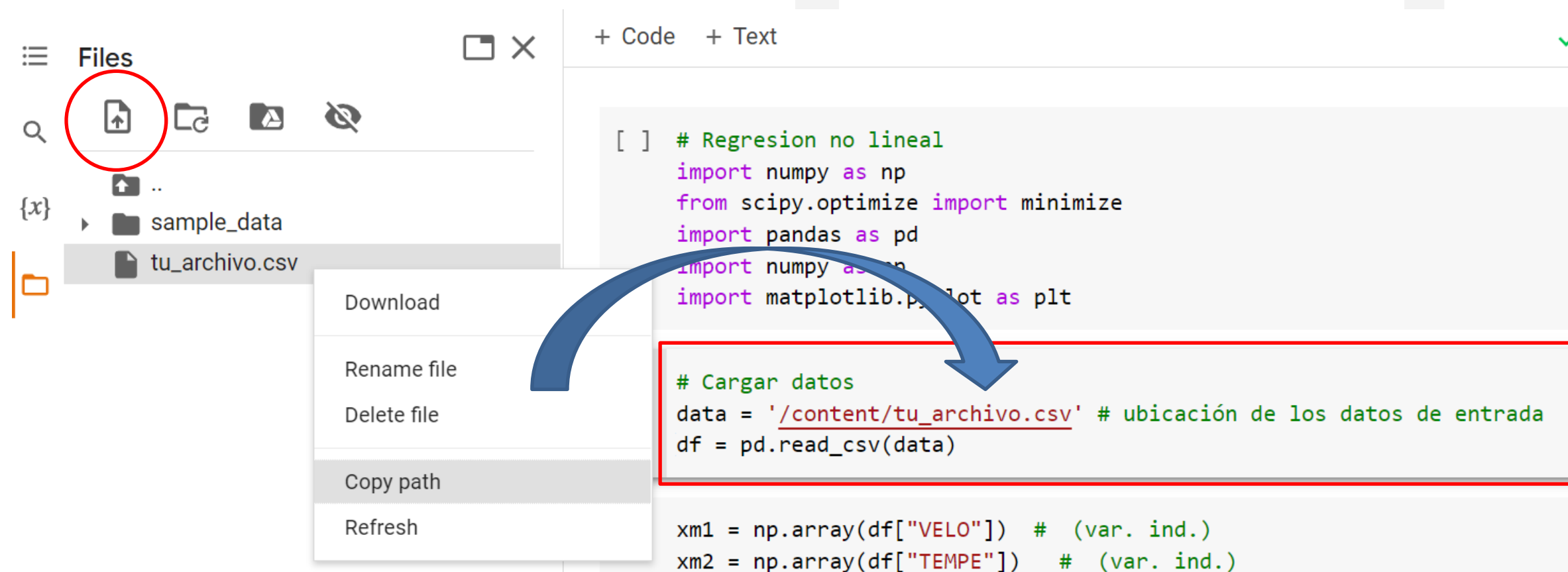
V. Ind.

V. Dep.

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	

## EJEMPLO DE APLICACIÓN

12. Ingresa a Google Colab (archivo “3\_Ejemplo\_4\_Visco\_Todas\_Temp.ipynb”)
13. Sube el archivo CSV del paso anterior.
14. Copia la ruta del archivo (/content/tu\_archivo.csv).
15. Pega la ruta en el bloque de código de carga de datos.

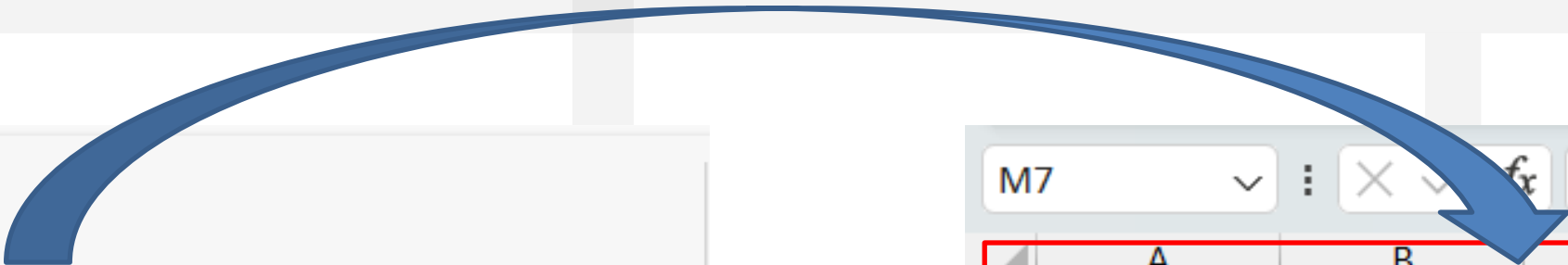




## EJEMPLO DE APLICACIÓN

- Después de la sección de carga de datos en el código, verifica que las cabeceras escritas en la declaración de arreglos correspondan con las escritas en el archivo CSV que se cargó.
- Escribe el modelo para el cual quieres realizar la regresión.

```
▶ 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
```



	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. Correr el programa dando *click* al símbolo de “*play*” en todos los bloques de código de manera descendente.

```
✓ 0s  # 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] ✓ 0s # Cargar datos
data = '/content/6_Ejemplo_4_Visco_Todas_Temp.xlsx.csv' # ubicación de los datos de entrada
df = pd.read_csv(data)

✓ 0s  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
```



## EJEMPLO DE APLICACIÓN

9. Al final del último bloque de código se pueden visualizar los resultados: coeficientes de regresión,  $R^2$  y gráfico de variable calculada vs variable real.

Solucion

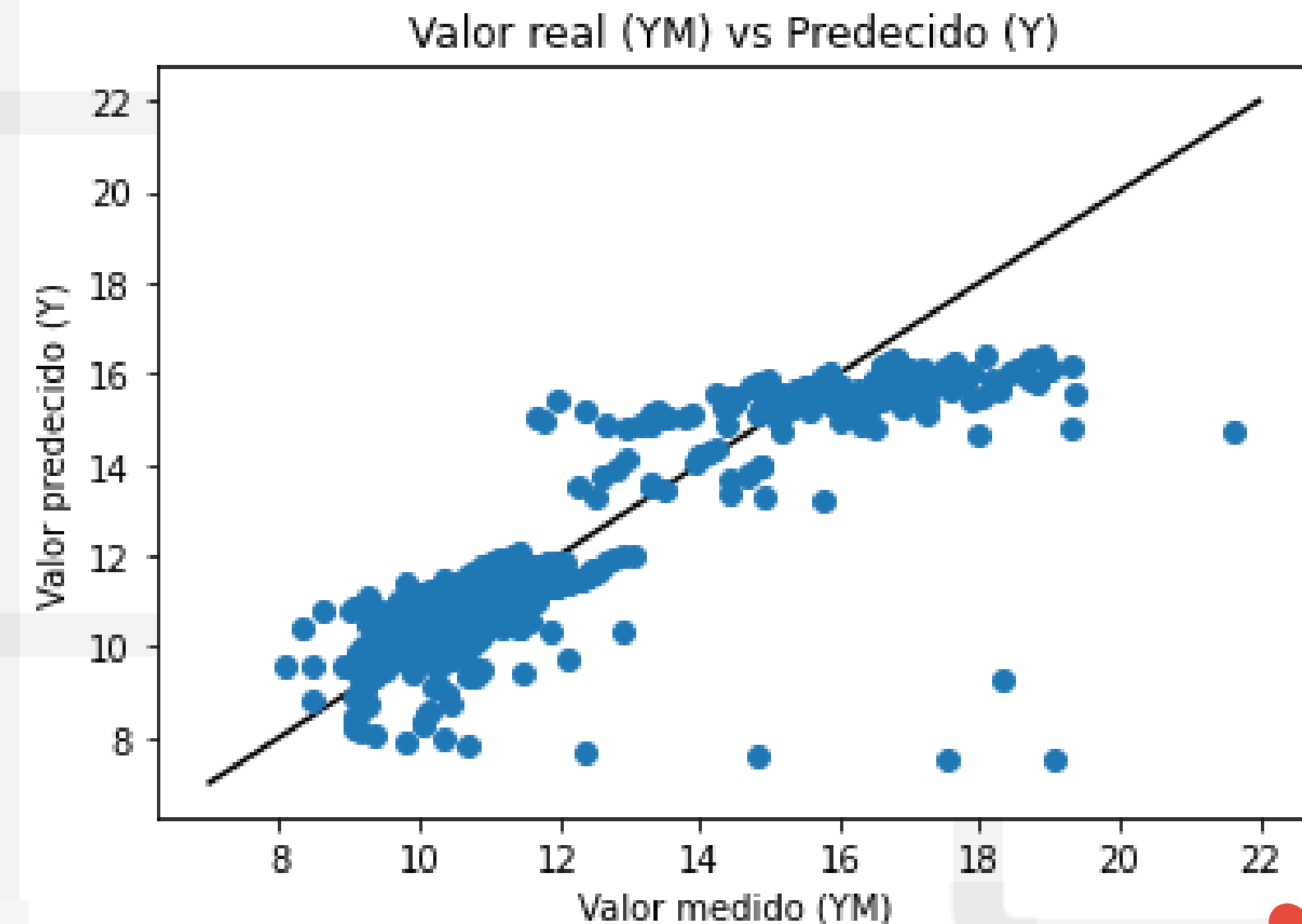
A = 19.074265749320542

B = 0.006369106592363837

C = -0.002168512964455647

D = 42.131902949314124

$R^2$  de correlacion = 0.7778296785295723

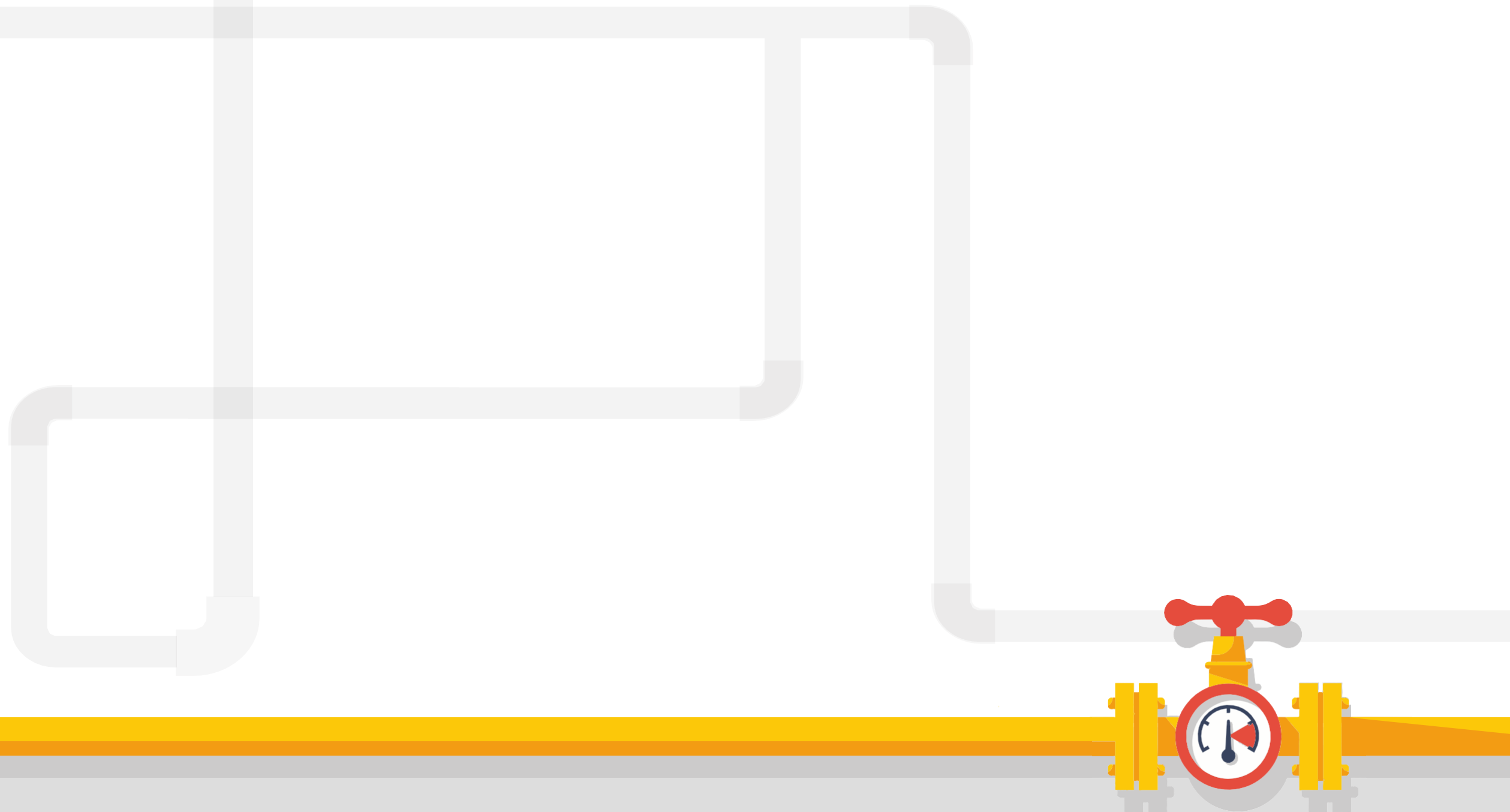


# EJERCICIO DE REFORZAMIENTO

---

## Quiz #4:

<https://acortar.link/O3NbDZ>





Universidad Nacional de Colombia  
SPE Student Chapter

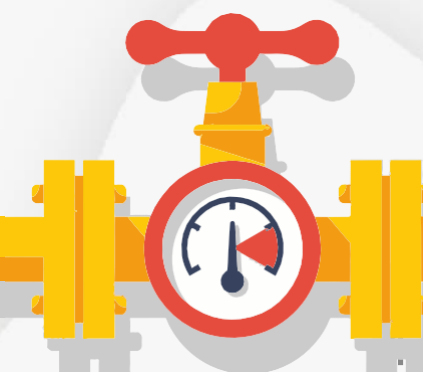
PETROIntelligence.com

## SESIÓN 5 REGRESIONES APLICADAS A LA INGENIERÍA DE PETRÓLEOS (CON PYTHON)

Misael Edgar Zepeda Díaz

[mzepeda@petrointelligence.com](mailto:mzepeda@petrointelligence.com)  
(+52 55 1350 7432)

Agosto 2023



Título	Horas
1. Introducción -----	1
2. Manejo de datos -----	1
3. Python -----	2
4. Regresiones -----	1
5. Análisis de los resultados -----	1
Total =	6

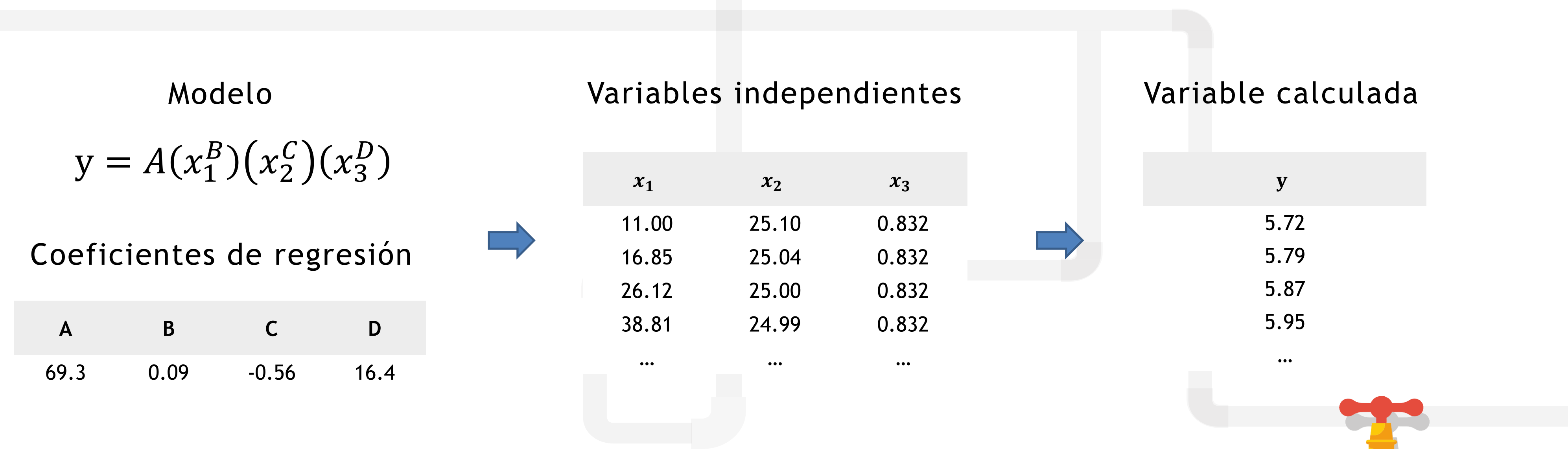


### Criterios de evaluación de la regresión:

1. Calcular los valores objetivo con los coeficientes de regresión y con el modelo propuesto.
2. Graficar los valores calculados vs los valores reales y la línea de referencia.
3. Calcular el error relativo promedio (ERP).
4. Calcular la desviación promedio (DP).
5. En caso necesario, evaluar una muestra de los datos para comparar los gráficos de los datos reales vs los datos estimados en función de alguna de las variables independientes.

Criterios de evaluación de la regresión:

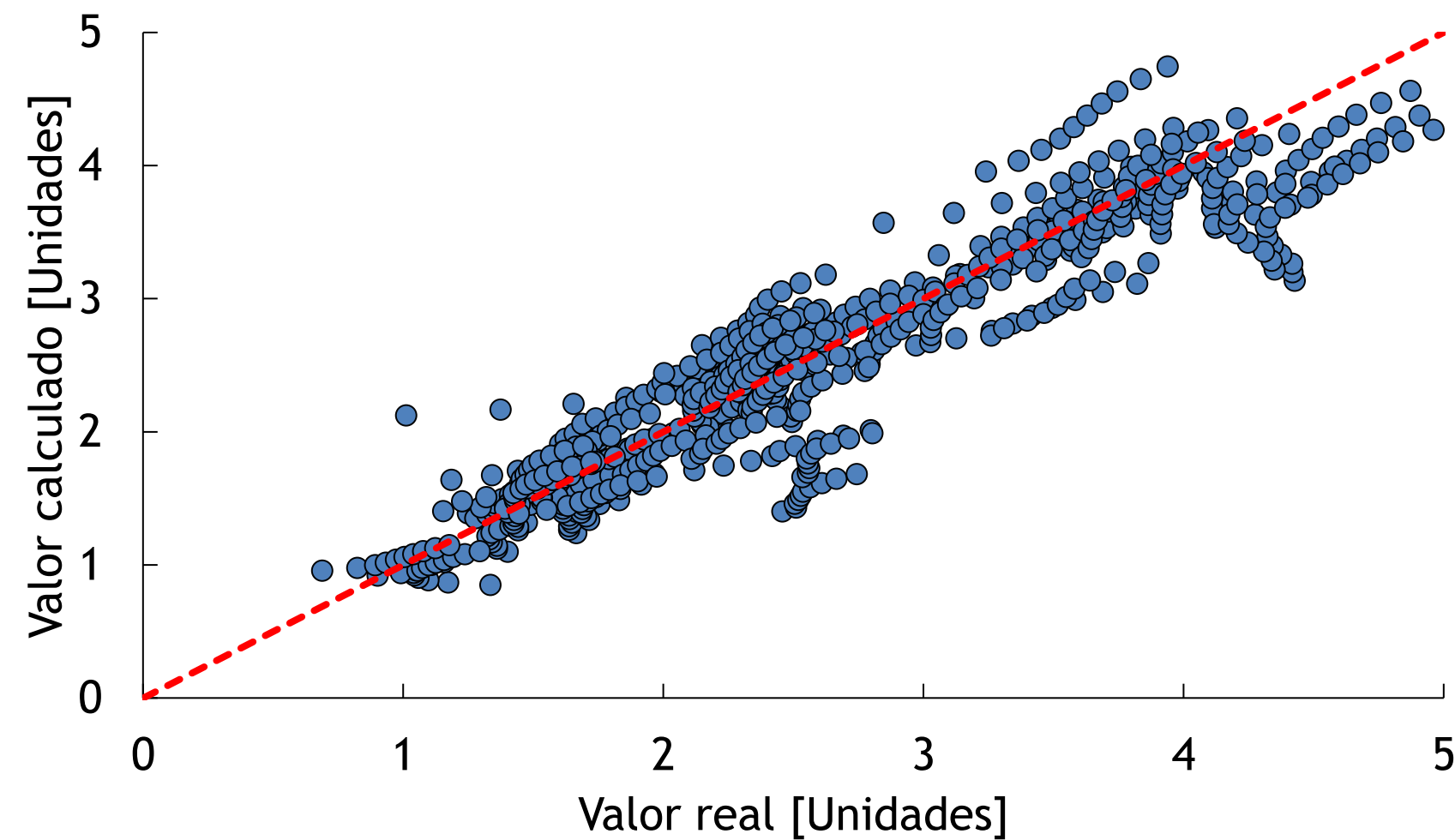
- 1. Calcular los valores objetivo con los coeficientes de regresión y con el modelo propuesto.





### Criterios de evaluación de la regresión:

2. Graficar los valores calculados vs los valores reales y la línea de referencia.



### Criterios de evaluación de la regresión:

#### 3. Calcular el error relativo promedio (ERP):

##### I. Calcular el error relativo de cada dato calculado con respecto al dato real:

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

##### II. Calcular el error relativo promedio:

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

### Criterios de evaluación de la regresión:

#### 4. Calcular la desviación promedio (DP):

##### I. Calcular la desviación de cada dato:

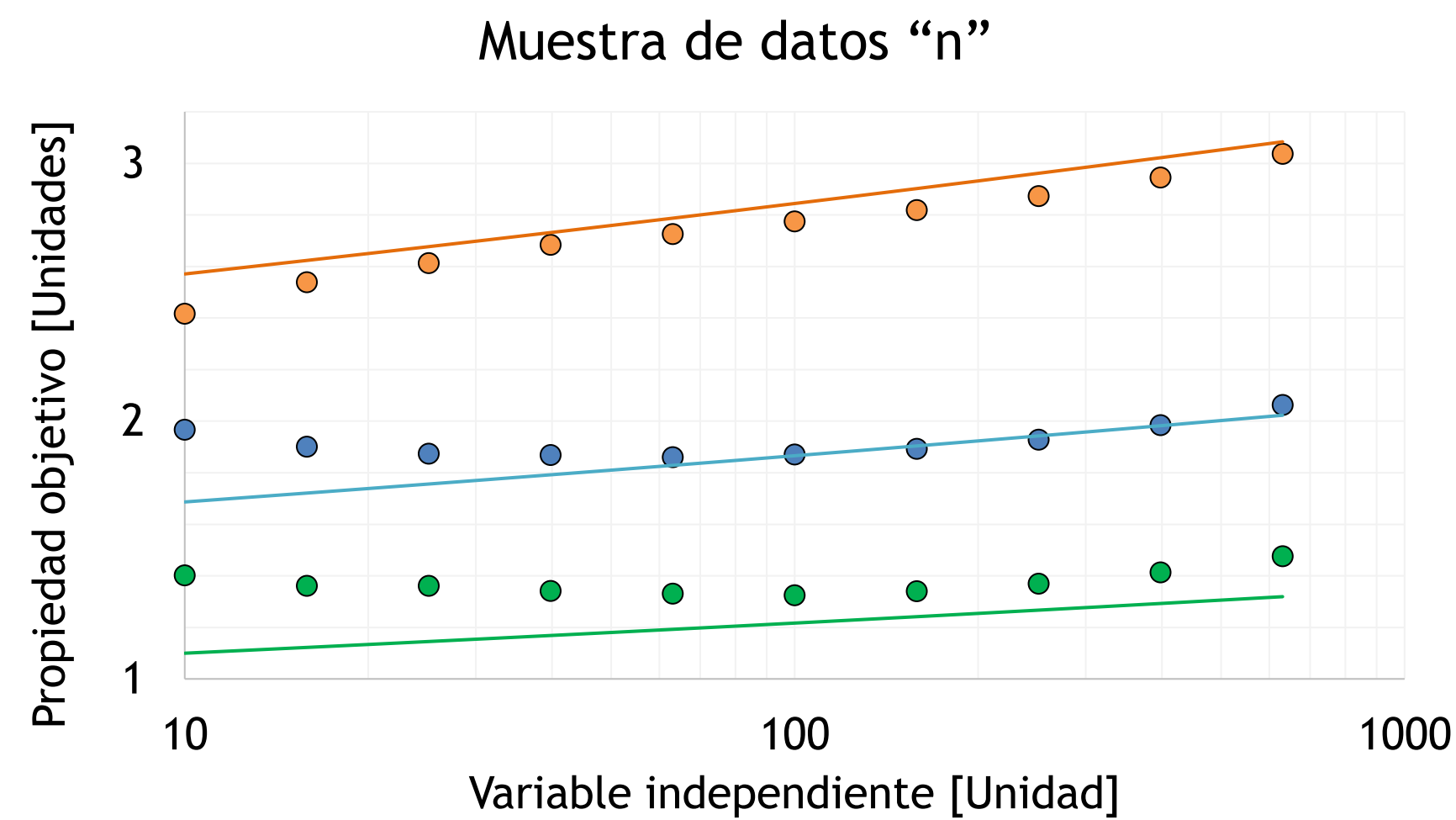
$$D = |Valor\ calculado - Valor\ real|[Unidades]$$

##### II. Calcular la desviación promedio:

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

Criterios de evaluación de la regresión:

5. En caso necesario, evaluar una muestra de los datos para comparar los gráficos de los datos reales vs los datos estimados en función de alguna de las variables independientes.



Las líneas continuas son las estimaciones.  
Los puntos son los datos reales.

Curva	Error relativo promedio %	Desviación promedio [Unidad]
Amarilla	2.92	0.079
Azul	3.92	0.075
Verde	11.99	0.165
Promedio =	6.28	0.11



### Ejemplo de análisis de los resultados para el Consumo de Gas Natural

1. Reorganizar la base de datos de consumo de gas natural.
2. Colocar los coeficientes de regresión.
3. Colocar las  $R^2$ .
4. Aplicar el modelo de regresión a los datos.
5. Calcular el error relativo promedio.
6. Calcular la desviación promedio.
7. Graficar resultados y valores reales vs la variable independiente.

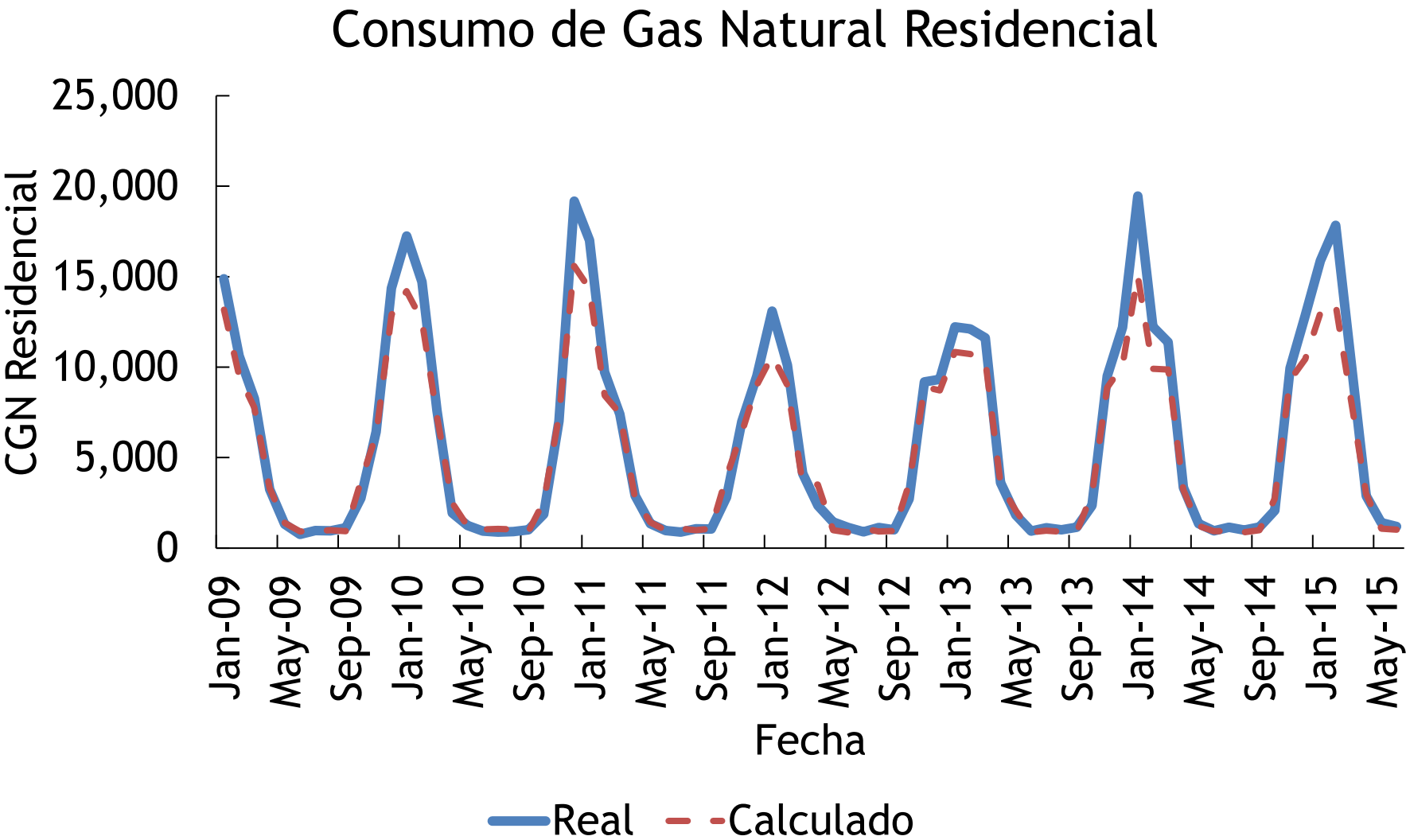
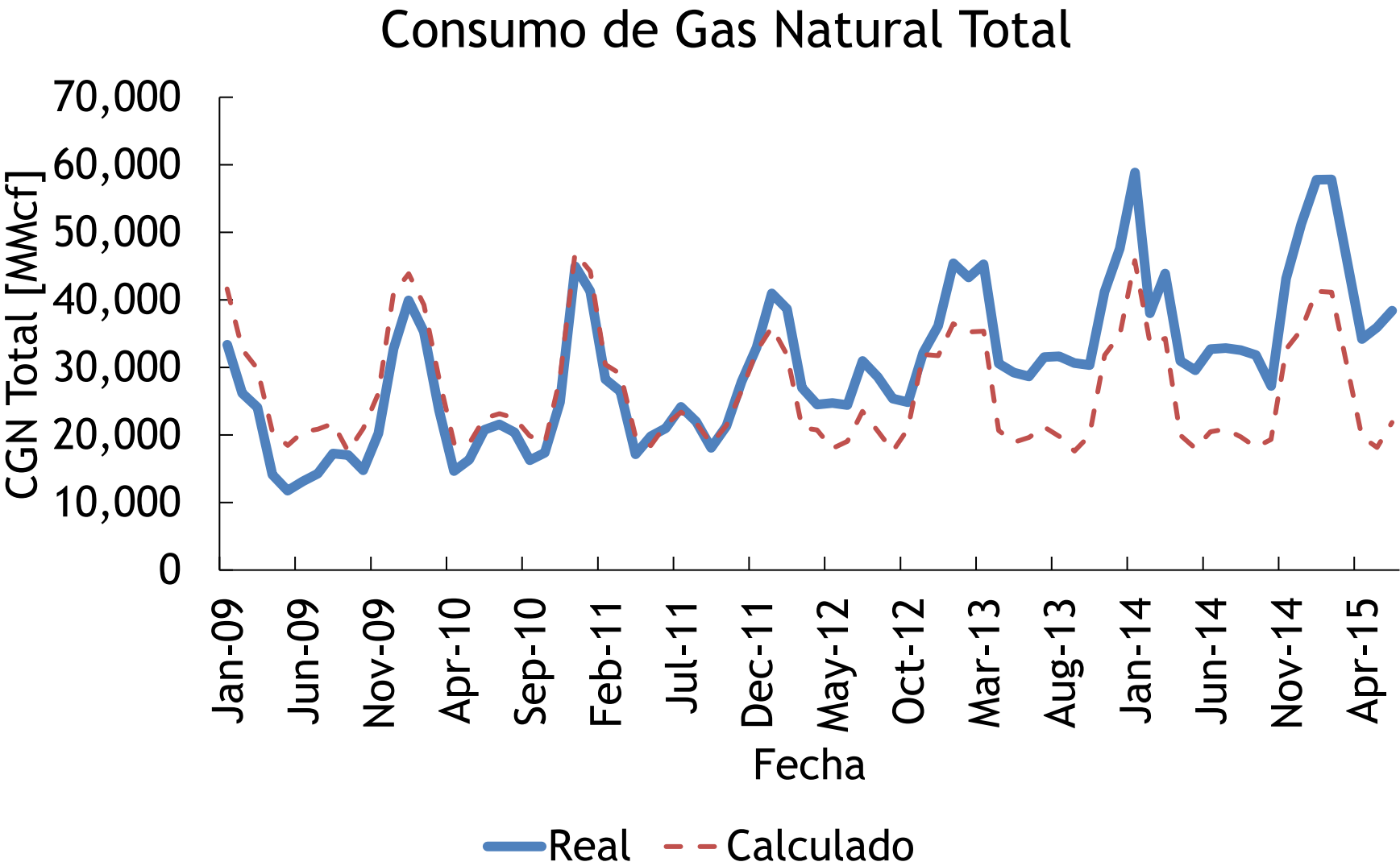
Ejemplo de análisis de los resultados para el Consumo de Gas Natural

Resumen de coeficientes de regresión y  $R^2$ :

			X1 = CCD	X2 = HDD	X3 = MT	Ind.	
Regresión	Var. Ind.	Modelo	A	B	C	D	$R^2$
1	CGN Total	LM	6.1914574	166.508598	911.617585	50.2949475	0.53488825
2	CGN Residencial	LM	-1.72152896	56.0627504	83.0634466	4.71463202	0.98624988
3	CGN Comercial	LM	-8.96678797	33.8539185	147.918148	8.20743716	0.96448859
4	CGN Industrial	LM	-28.8382881	31.5547293	423.909496	23.3505471	0.47640213
5	CGN Eléctrico	LM	49.41608	17.3082204	120.169503	6.49220244	0.09443807



Ejemplo de análisis de los resultados para el Consumo de Gas Natural



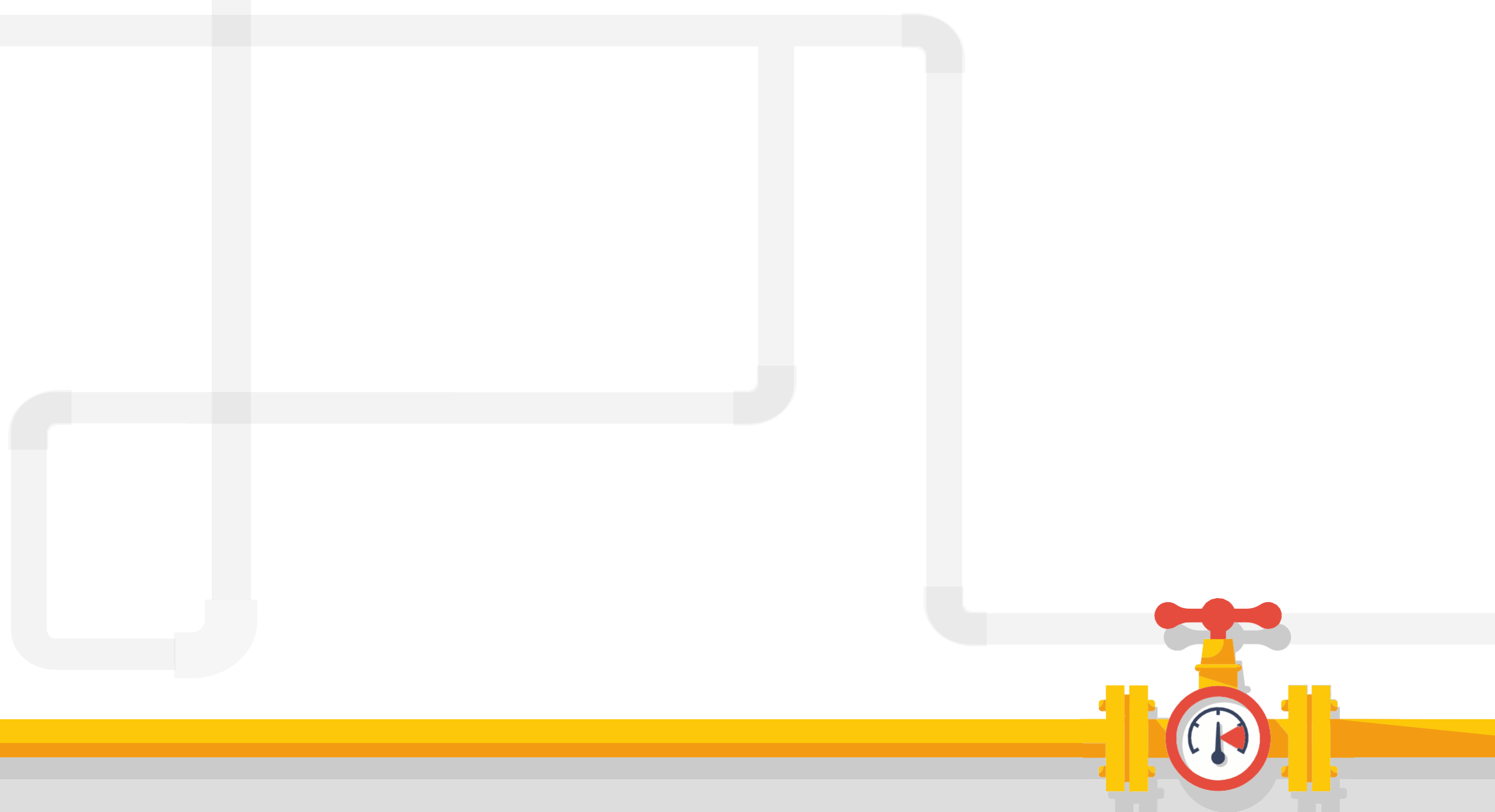


# EJERCICIO DE REFORZAMIENTO

---

## Quiz #5:

<https://acortar.link/04snwK>



## Quizzes:

1. <https://acortar.link/i9wtPG>
2. <https://acortar.link/tganuo>
3. <https://acortar.link/Ok7IIQ>
4. <https://acortar.link/O3NbDZ>
5. <https://acortar.link/04snwK>





Universidad Nacional de Colombia  
SPE Student Chapter

**PI** PETRO  
Intelligence®

