

```
!pip install pandas
!pip install scikit-learn
!pip install matplotlib
!pip install scipy
!pip install --upgrade pandas lxml html5lib
```

```
Requirement already satisfied: pandas in
/usr/local/lib/python3.10/dist-packages (2.2.2)
Requirement already satisfied: numpy>=1.22.4 in
/usr/local/lib/python3.10/dist-packages (from pandas) (1.25.2)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2023.4)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2024.1)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2-
>pandas) (1.16.0)
Requirement already satisfied: scikit-learn in
/usr/local/lib/python3.10/dist-packages (1.2.2)
Requirement already satisfied: numpy>=1.17.3 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.25.2)
Requirement already satisfied: scipy>=1.3.2 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.11.4)
Requirement already satisfied: joblib>=1.1.1 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.5.0)
Requirement already satisfied: matplotlib in
/usr/local/lib/python3.10/dist-packages (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (1.2.1)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (4.53.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (1.4.5)
Requirement already satisfied: numpy>=1.20 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (1.25.2)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (24.1)
Requirement already satisfied: pillow>=6.2.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in
```

```
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7-
>matplotlib) (1.16.0)
Requirement already satisfied: scipy in
/usr/local/lib/python3.10/dist-packages (1.11.4)
Requirement already satisfied: numpy<1.28.0,>=1.21.6 in
/usr/local/lib/python3.10/dist-packages (from scipy) (1.25.2)
Requirement already satisfied: pandas in
/usr/local/lib/python3.10/dist-packages (2.2.2)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-
packages (5.2.2)
Requirement already satisfied: html5lib in
/usr/local/lib/python3.10/dist-packages (1.1)
Requirement already satisfied: numpy>=1.22.4 in
/usr/local/lib/python3.10/dist-packages (from pandas) (1.25.2)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2023.4)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2024.1)
Requirement already satisfied: six>=1.9 in
/usr/local/lib/python3.10/dist-packages (from html5lib) (1.16.0)
Requirement already satisfied: webencodings in
/usr/local/lib/python3.10/dist-packages (from html5lib) (0.5.1)
```

```
import math
import pandas as pd
from sklearn import linear_model
from sklearn.metrics import mean_absolute_error
import matplotlib.pyplot as plt
from scipy.stats import pearsonr

# 0. Cargar los datos de estatura y peso
url = 'http://wiki.stat.ucla.edu/socr/index.php?
title=SOCR_Data_Dinov_020108_HeightsWeights&oldid=12529'

height_weight_df = pd.read_html(url)[1]
[['Height(Inches)', 'Weight(Pounds)']]

# 1. Contar registros en el dataframe
record_count = len(height_weight_df)
print(f'Número de registros en el dataframe: {record_count}')
```

Número de registros en el dataframe: 200

```
# 2. Crear una regresion lineal con estatura como variable
explicatoria y peso como variable explicada
X = height_weight_df[['Height(Inches)']]
y = height_weight_df['Weight(Pounds)']
```

```
reg = linear_model.LinearRegression()  
reg.fit(X, y)
```

```
LinearRegression()
```

```
# 3. Hallar el intercepto
```

```
intercept = reg.intercept_
```

```
slope = reg.coef_[0]
```

```
print(f'Intercepto: {intercept}')
```

```
print(f'Pendiente: {slope}')
```

```
Intercepto: -106.02770644878137
```

```
Pendiente: 3.4326761292716297
```

```
# 4. Hacer la gráfica
```

```
plt.scatter(X, y, color='blue')
```

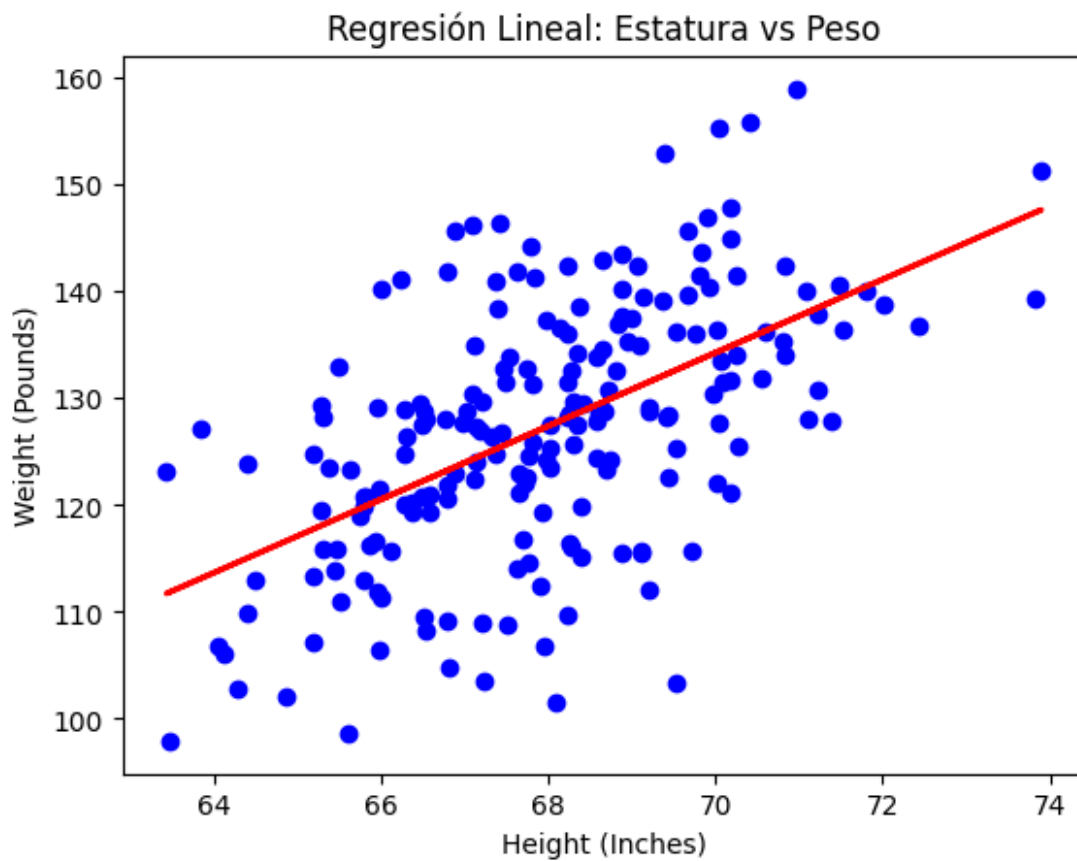
```
plt.plot(X, reg.predict(X), color='red', linewidth=2)
```

```
plt.xlabel('Height (Inches)')
```

```
plt.ylabel('Weight (Pounds)')
```

```
plt.title('Regresión Lineal: Estatura vs Peso')
```

```
plt.show()
```



```
# 5. Hallar correlación y valor p
correlation, p_value = pearsonr(height_weight_df['Height(Inches)'],
height_weight_df['Weight(Pounds)'])
print(f'Correlación: {correlation}')
print(f'Valor p: {p_value}')
```

```
Correlación: 0.5568647346122995
Valor p: 1.1029015151265851e-17
```

```
# 6. Interpretar el valor p
interpretation = ""
if p_value < 0.05:
    interpretation = "La relación entre estatura y peso es
estadísticamente significativa."
else:
    interpretation = "La relación entre estatura y peso no es
estadísticamente significativa."

print(interpretation)
```

```
La relación entre estatura y peso es estadísticamente significativa.
```