

Linear Regression: Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It is used for prediction and data analysis.

The equation for simple linear regression is: $y = mx + c$.

y: Dependent variable, x: independent variable, c: constant, m: slope

$$\text{Correlation Coefficient } r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$\bar{x} = \text{mean of variable } X = \frac{\sum x}{n}$$

$$\bar{y} = \text{mean of variable } Y = \frac{\sum y}{n}$$

slope $m = r \times (S_y / S_x)$, where S_x is standard deviation of x and S_y is standard deviation of y.

constant $c = \bar{y} - \bar{x} * m$

Refer the 17-Regression.xlsx file for example.

```
import matplotlib.pyplot as plt
from scipy import stats

x = [2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011,
2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023,
2024]
y = [3262.33, 3377.28, 5838.96, 6602.69, 9397.93, 13786.91, 20286.99,
9647.31, 17464.81, 20509.09, 15454.92, 19426.71, 21170.68, 27499.42,
26117.54, 26626.46, 34056.83, 36068.33, 41253.74, 47751.33, 58253.82,
60840.74, 72240.26, 78139.01]

slope, intercept, r, p, std_err = stats.linregress(x, y)

def myfunc(x):
    return slope * x + intercept

mymodel = list(map(myfunc, x))

plt.scatter(x, y)
plt.plot(x, mymodel)
plt.show()
```

