

```
In [ ]: from sklearn.datasets import load_boston
        from sklearn.linear_model import LinearRegression
        import matplotlib.pyplot as plt
        import pandas as pd
        import numpy as np
```

Διάνυσμα με 256 στοιχεία με τιμές από το 0 μέχρι το 100 (από ομοιόμορφη κατανομή)

```
In [ ]: x = 100 * np.random.rand(256)
```

Δημιουργία των δεδομένων χρησιμοποιώντας τη διαδικασία
$$y = A + Bx + \epsilon$$

```
In [ ]: A = 10
        B = 4
        sigma_epsilon = 10
```

```
In [ ]: y = B * x + sigma_epsilon * np.random.randn(256) + A
```

```
In [ ]: dataset = pd.DataFrame()
```

```
In [ ]: dataset['x'] = x
        dataset['y'] = y
```

```
In [ ]: dataset.plot.scatter(x='x', y='y')
```

Χρήση της υλοποίησης του μοντέλου γραμμικής παλινδρόμησης από sklearn

```
In [ ]: reg = LinearRegression()
```

Προσαρμογή της ευθείας στις παρατηρήσεις (συνάρτηση fit)

```
In [ ]: reg.fit(np.expand_dims(x,1),y)
```

Υπολογισμός του a , είναι $a = \hat{y}(0)$

```
In [ ]: a = reg.predict(np.expand_dims(np.array([0]), -1))[0]
```

Η παράμετρος της κλίσης b περιέχεται στο

```
In [ ]: b = reg.coef_[0]
```

```
In [ ]: a
```

```
In [ ]: b
```

Υπολογισμός του $\hat{y}(100)$ ώστε να έχουμε 2ο σημείο (το πρώτο είναι το $(0, a)$) για να σχεδιάσουμε την ευθεία

```
In [ ]: y_hat_100 = reg.predict(np.expand_dims(np.array([100]), -1))[0]
```

```
In [ ]: dataset.plot.scatter(x='x', y='y')
plt.plot((0,100),(a, y_hat_100), color = 'red', lw=3)
```

Άσκηση : Υπολογίστε το συντελεστή προσδιορισμού R^2

Άσκηση : Δοκιμάστε άλλες τιμές των A,B και sigma_epsilon

```
In [ ]:
```