

Simple Linear Regression with Stats Models

We seek to solve two problems.

① Linear Regression with uncertainties in the fit parameters.

statsmodels →

② A nicer data format that does not require 2D matrices, etc.

pandas →

Step 1 : Get the data into a data structure.

```
import pandas as pd
```

```
df = pd.DataFrame(  
    [ 1, 2, 3, 4, 5, 10, 15, ... ],
```

$\{ x : [5, 12, 20, \dots] \}$
column labels
data for each column

→ Basically, it's creating a big Excel Spreadsheet, with column headings, ...

`print(df)`
`print(df['x'])`

refer to a specific column.

Step 3 : Fit the data

`y = df['y']`

$$X = df['x']$$

$$X = sm.add_constant(X)$$

add a "constant"
term to the fit.

$$y = \beta_0 + \beta_1 X$$

$$\text{model} = sm.OLS(y, X).fit()$$

Note order!!!

"Ordinary Least Squares"
≡ Basic Linear Regression

print(model.summary())

the best thing ever!!!