

Regression Plots

In this lesson, regression plots with the seaborn package are discussed.

We'll cover the following ^

- Introduction
- Subset of regression

Introduction

Regression is usually used to find the relationship between the dependent and independent variables of a dataset. Similarly, a regression plot helps us visualize the relationship of data in real time. By successfully obtaining a regression line, analysts can use it to predict future trends of the data. Let's learn how to make a regression plot.

In the seaborn package, the `lmplot` function is used to generate regression plots. Seaborn also provides built-in datasets to test its functionalities.

We will use the `tips` dataset as used in its [documentation](#). The data information about customers at a restaurant.

```
import numpy as np
import pandas as pd
import seaborn as sns

df = sns.load_dataset('tips') # Loading dataset from seaborn library

sns1 = sns.lmplot(x = 'total_bill', y = 'tip', data = df) # plotting regression plot
```



In the above example, the `lmplot()` function takes the `x` value as the independent variable and the `y` value as the dependent variable to calculate the regression line, as discussed above. The `total_bill` and `tip` are two attributes of the data set we are using. As the tip is calculated after the total bill is generated, the `tip` variable is dependent on the `total_bill` variable of the `tips` dataset.

It can be seen in the output that a *regression* line is drawn. The shaded area around the line indicates that the true regression line lies within the shaded region. This line indicates that the greater the bill, the more the tip amount, i.e., an increase in the bill results in an increase in the tip as well. This same line is used in statistics to generate an equation describing the relationship between the variables.

Subset of regression

We can use other attributes of the dataset to plot a subset regression plot with the original regression plot. It helps find relationships of different attributes with each other. Let's understand this with the following code example.

```
import numpy as np
import pandas as pd
import seaborn as sns

df = sns.load_dataset('tips')

sns1 = sns.lmplot(x = 'total_bill', y = 'tip', hue = 'sex', data = df)
```

If the plot in the output is zoomed in, we can see that two regression lines are generated for men and women separately. Information, such as which *gender* tips more, can be analyzed from it. Similarly, it allows us to predict the behavior of the dependent variables with respect to a number of independent variables.

For more functionalities and information on regression plots, refer [here](#).

In the next lesson, heatmaps are discussed.