STAT 107 – Spring 2021

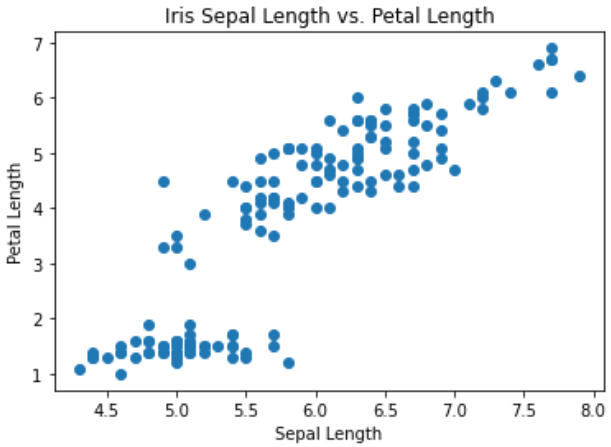
Measuring a Linear Relationship

Ha Khanh Nguyen (hknguyen)

Review: Scatterplots

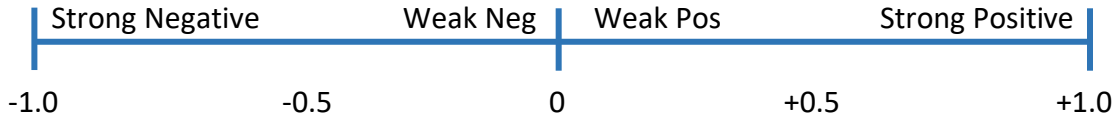
Two methods:

```
import matplotlib.pyplot as plt
import pandas as pd

plt.scatter(x, y)

df.plot.scatter(x, y)
#where df is the name of the DataFrame
containing the variables
```



Correlation Coefficient

- It measures the strength of the linear association between two variables (X and Y).
- The sample correlation coefficient is notated as r where the population correlation coefficient is denoted as ρ.
- Correlation coefficient can take any value between -1 and 1.
- Negative values imply that as one variable increases in value, the other decreases in value.
- Positive values imply that as one variable increases in value, the other increases in value as well.



Computing Correlation Coefficient

There are a number of correlation coefficient formulas in Statistics. The one we will be using is called Pearson's correlation coefficient:

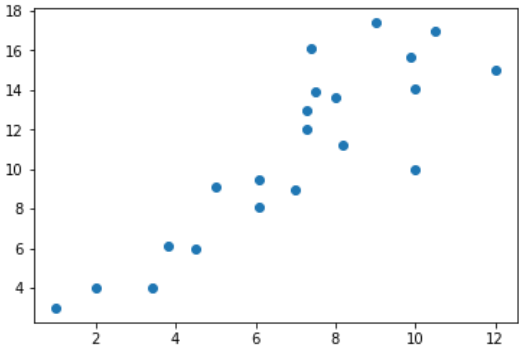
$$r = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

Luckily, we don't have to compute this by hand! We can use Python to compute r for us!

```
# method 1
x.corr(y)# x and y MUST be columns of a DataFrame or Series

# method 2
from scipy.stats import pearsonr
pearsonr(x, y)  # x and y can be lists, arrays, etc.
```

Correlation Coefficient Properties

```
x = [1, 2, 3.4, 8, 10.5, 9, 4.5, 5, 7, 6.1, 3.8,
6.1, 7.4, 8.2, 9.9, 10, 12, 7.3, 7.5, 7.3, 10]
y = [3, 4, 4, 13.6, 17, 17.4, 6, 9.1, 9, 8.1, 6.1,
9.5, 16.1, 11.2, 15.7, 14.1, 15, 12, 13.9, 13, 10]

plt.scatter(x, y)
pearsonr(x, y)
```



Let's try to change X and Y values and see how that affects the correlation coefficient r of X and Y.

- Adding 10 to every X value.
- Switching Y values with the 3 lowest values with the 3 highest X values.
- Switching X and Y.
- Multiplying all of the X values by 2.
- Multiplying all of the X values by -5.

Matching Scatterplots with Correlation Coefficients

$r = 0.65$
$r = 0.98$
$r = -0.84$
$r = -0.16$
$r = 0.9$
$r = 0.62$