

Statistics Cheat Sheet- 3

Covariance

- Measures the direction of the relationship between variables, and how much one variable depends on the other.
- A positive covariance means that both variables tend to be high or low at the same time.
- A negative covariance means that when one variable is high, the other tends to be low.
- A covariance of zero indicates that there is no clear directional relationship between the variables being measured.
- Covariance values can range from negative infinity to positive infinity, and are affected by changes in scale.

```
import numpy as np
# Covariance Calculation
cov_matrix = np.cov(X, Y)
# Covariance between X and Y
cov_XY = cov_matrix[0, 1]
# Covariance Matrix
cov_matrix = df.cov()
```

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

For a sample covariance, the formula is slightly adjusted:

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

Correlation

- While covariance measures the direction of a relationship between two variables, correlation measures the strength of that relationship.
 - Correlation, $\text{Corr}(X, Y)$, scales covariance by the standard deviations of X & Y.
 - Correlation values range from -1 to +1, and are not affected by changes in scale.
 - Covariance is limited to measuring the relationship between two variables, while correlation can be used for multiple sets of numbers.
- $r=1$: Perfect positive correlation.
 - $r=-1$: Perfect negative correlation.
 - $r=0$: No correlation.

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

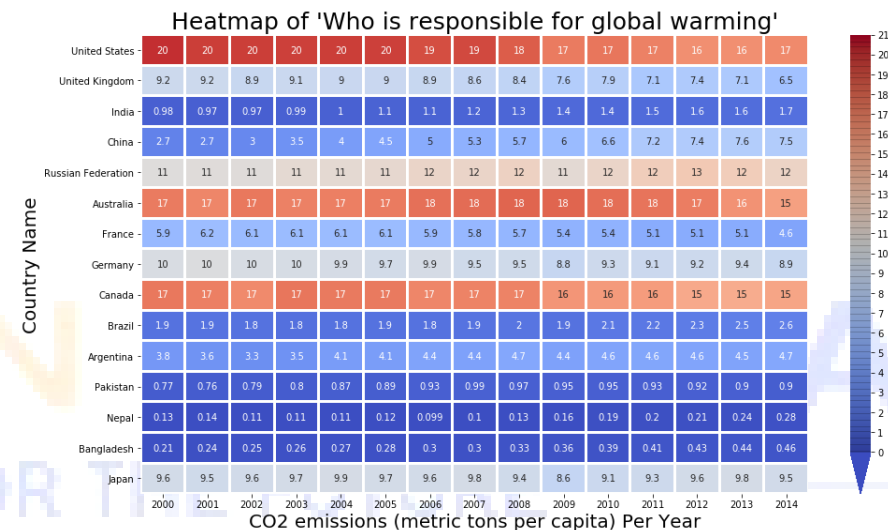
```
import numpy as np
# Correlation Calculation
corr_matrix = np.corrcoef(X, Y)
# Correlation between X and Y
corr_XY = corr_matrix[0, 1]
# Correlation Matrix
corr_matrix = df.corr()
```

Correlation Heatmap

A correlation heatmap is a visual representation of the correlation matrix, which shows the pairwise correlation between features.

Focus on features with low correlations among themselves but strong correlations with the target variable.

```
import seaborn as sns
import matplotlib.pyplot as plt
# Heatmap of Correlation Matrix
sns.heatmap(corr_matrix, annot=True, cmap="coolwarm", fmt=".2f")
```

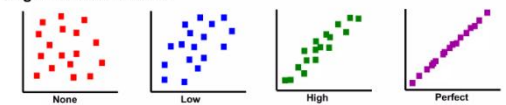


Scatter Diagram

A scatter diagram is a tool for analyzing relationships between two variables for determining how closely the two variables are related.

```
import matplotlib.pyplot as plt
plt.scatter(x, y, color='blue', alpha=0.7)
# Pair Plot
sns.pairplot(df, diag_kind='hist')
plt.show()
```

Degrees of correlation:



Types of correlation:

