공분산이라 하면 분산과는 다르게
하나의 변수가 아닌 두 변수 사이의 관계를 나타낸다고 생각하면 될 것이다.
체고와 도체중의 관계가 궁금하다고 가정하자.
체고가 커질수록 도체중이 증가할까?
물론 당연히 도체중이야 증가한다.
도체수율은 부의상관을 나타내지만... (상관 관계)

각설하고, 이 두 변수간의 변동을 공분산 Cov(X, Y) 이라 한다.

공분산 값은 아래와 같이 나타낸다.

Cov(X, Y) > 0 X가 증가 할 때 Y도 증가한다.
Cov(X, Y) < 0 X가 증가 할 때 Y는 감소한다.
Cov(X, Y) = 0 공분산이 0이라면 두 변수간에는 아무런 선형관계가 없으며 두 변수는 서로 독립적인 관계에 있음을 알 수 있다.
그러나 두 변수가 독립적이라면 공분산은 0이 되지만, 공분산이 0이라고 해서 항상 독립적이라고 할 수 없다.

공분산의 개념은 우리가 흔히 사용하는 상관계수와 연관지어 생각해 보아야 한다.
공분산을 구하다 보면,
공분산 값이 항상 일정하지 않기 때문에 비교하고자 한다면 계산도 해야하며 머리가 아파온다.

-000000... <= Cov(X, Y) <= 000000...(ex)

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

그래서 이를 표준화 시켜주는 작업으로 "공분산에 표준편차로 나누어 주면
값이 -1 <= Corr(x, y) <= 1 사이 범위로 좁혀지면서 우리는 쉽게 비교할수가 있어진다.
이것이 바로 상관계수 Corr(x, y)인 것이다.

X 공분산 행렬 데이터 매트릭스.

$$cov(X) = \frac{X^T \cdot X}{n}$$

cov(X): ↑ 대칭 행렬
n: ↑ or 'n-1'

- covariance matrix의 i번째 행과 j번째 열은 i번째 feature와 j번째 feature가 서로 함께 변하는 정도를 의미한다.

X linear correlation은 두 개의 양적 변수의 상관관계(선형 only)