

* 그래서 중요한 **연관성의 정도**

* **상관계수** (Correlation coefficient)

공분산을 표준화시켜주는 값을 상관계수라 부르고, 이것으로 연관성의 측정하도록 한 것임. 상관계수 서로 관계가 있다 해서 correlation coefficient (계수) 라 함.

- 상관계수 - X, Y의 공분산을 각각의 편차로 나누어 준다
공분산을 표준화한 것은 공분산

$$\text{Corr}(X, Y) = \frac{\text{COV}(X, Y)}{\sigma_X \sigma_Y} \rightarrow \text{공분산}$$

상관계수 범위에 해당되는 편차로 나눈 것 ←

$$\text{COV}(X, Y) = E(X - \mu_X)(Y - \mu_Y)$$

시그마 X $\sigma_X = \sqrt{\text{Var}(X)}$, $\sigma_Y = \sqrt{\text{Var}(Y)}$

(모집단 X의 표준편차) 표준편차 표준편차

* 피어슨 상관계수: r

표본에서 어떻게 구할 것인가?

• 표본 상관계수 → X에서 X 평균 빼고 → Y에서 Y 평균 빼고

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

μ \bar{X}
모집단의 평균 (무) 샘플의 표본평균 (X bar)
X의 평균

• Recall

$$\text{Corr}(X, Y) = \frac{E(X - \mu_X)(Y - \mu_Y)}{\sqrt{E(X - \mu_X)^2 E(Y - \mu_Y)^2}}$$

Correlation