2020년 10월 8일 목요일 오후 2:45

· 원화. · GDP

· 달러. · GDP

데이터의 분포를 비교하기 위해 특정 단위로 통일 → 정규화, 표준화.

⇒ 표준편차. 분산. ⇒ 데이터의 분포.

↓

데이터를 비교.

⇓

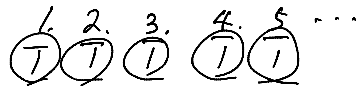
상관계수

정규화. : 데이터 → 0~1

$$X' = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

표준화 : $\frac{X - \bar{X}}{\sigma}$: $\bar{X}$: 평균

σ ← 「평균과의 거리」



$x_1 \quad x_2 \quad x_3 \quad x_4 \quad \cdots \quad x_N$

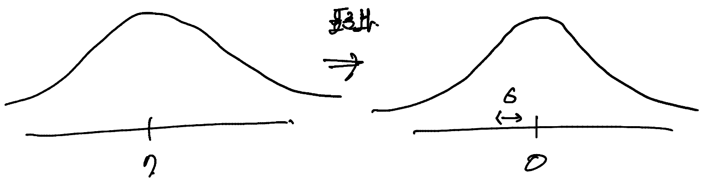
평균 : $\frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} = \bar{x}$ ⇒ $x_1 + \cdots + x_N = N \cdot \bar{x}$

분산. $V = s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2}{N}$ ⇒ $\sigma = \sqrt{s^2} = s$

$$\frac{x_1 - \bar{x} + x_2 - \bar{x} + \cdots + x_N - \bar{x}}{N}$$

$N \cdot \bar{x}$

$$\Rightarrow \frac{x_1 + \cdots + x_N - N\bar{x}}{N} = 0$$



공분산. $DataA_{cov}$ $DataB_{cov}$

공분산. DataA cov DataB cov

Corr.(DataA cov, DataB cov) : A와B의 상관계수

상관계수.