공분산(Covariance)과 상관계수(Correlation Coefficient)의 정의와 차이점

practice_선형대수 · 2024. 9. 29. 13:37

•공분산(Covariance)과 상관계수(Correlation Coefficient)의 정의와 차이점에 대 해 설명해 주세요.

공분산

•두개의 확률변수 X와 Y가 있을때
•공분산은 이 두 변수들이 얼마나 같이 움직이는지를 나타냄
•만약 공분산이 양수라면, 하나의 변수가 증가할 때 다른 변수 또한 증가하는 경향을 보임
•반대로 공분산이 음수라면, 하나의 변수가 증가할 때 다른 변수는 감소하는 경향을 보임

상관계수(Correlation Coefficient)

공분산을 표준화하여 두 변수 간의 선형 관계를 1과 -1 사이의 값으로 나타냅니다.

•상관계수는 공분산보다 해석이 용이합니다.

공분산과 상관계수의 차이점

확률변수 x, y에 대하여

공분산: 근본적인 비례 관계가 동일하더라도, 단위가 달라지면 값도 달라진다 (단점)

상관계수: 단위가 달라도 변하지않는다.(단점극복)- 근본적인 비례 관계만 고려한다.

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Y의 표준편차

X의 표준편차

공분산

•<u>두개의 확률변수 X와 Y가 있을때</u>

•<u>공분산은 이 두 변수들이 얼마나 같이 움직이는지를 나타냄</u>

•만약 공분산이 양수라면, 하나의 변수가 증가할 때 다른 변수 또한 증가하는 경향을 보임

•반대로 공분산이 음수라면, 하나의 변수가 증가할 때 다른 변수는 감소하는 경향을 보임

데이터셋 1

집값과 평수가 같이 증가

→ 공분산 양수

	집값	평수
1	70	60
2	120	115
3	150	130

데이터셋 2

집값과 평수간의 큰 관계 없음

→ 공분산이 0에 가깝다.

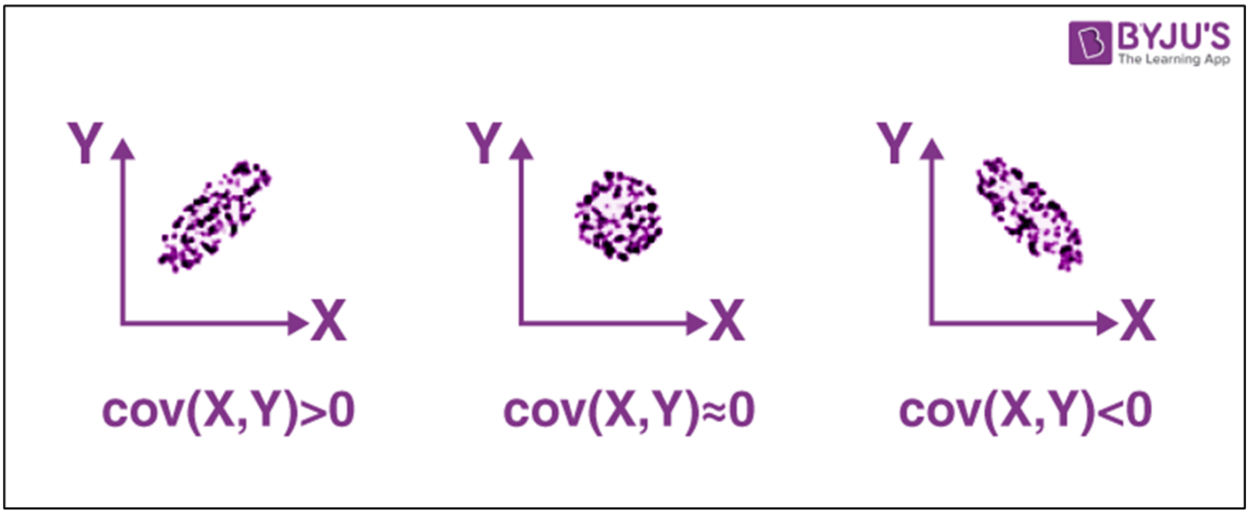
	집값	평수
1	120	60
2	130	125 ?
3	150	80

데이터셋 3

집값과 평수가 반대로 움직임

→ 공분산 음수

	집값	평수
1	70	135
2	130	110
3	150	80



아래그림) n-1로 나눈것은 n으로 나눈것보다 정확해서.

집1 집2 → N번째 집
평수 가격 평수 가격

- 수학적으로 표현하자면,
- X와 Y 값으로 구성된 데이터에 대하여 (x1, y1), (x2, y2), ... , (xN, yN)

- X와 Y의 공분산은 아래와 같이 계산된다

$$\boxed{cov_{x,y} = \frac{\sum(\overset{150}{x_i} - \overset{100}{\bar{x}})(\overset{130}{y_i} - \overset{100}{\bar{y}})}{N-1}}$$

- x̄ 와 ȳ 는 각각 X와 Y의 평균

아래그림)

근본적인 비례관계는 동일하지만,

공분산이 사용하는 단위가 달라지면 공분산도 달라진다.

집 넓이 (m^2)	집 가격
100	100
200	200
300	300

= 양수의 공분산 (어떤 값)

$3.3\,m^2 = 1$평

집 넓이 (평)	집 가격
$100 \div 3.3$	100
$200 \div 3.3$	200
$300 \div 3.3$	300

→ 본질적인 비례관계는 동일하지만, 사용하는 단위에 따라 공분산이 달라진다

$$\boxed{cov_{x,y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N-1}}$$

상관계수(Correlation Coefficient)

공분산을 표준화하여 두 변수 간의 선형 관계를 1과 -1 사이의 값으로 나타냅니다.

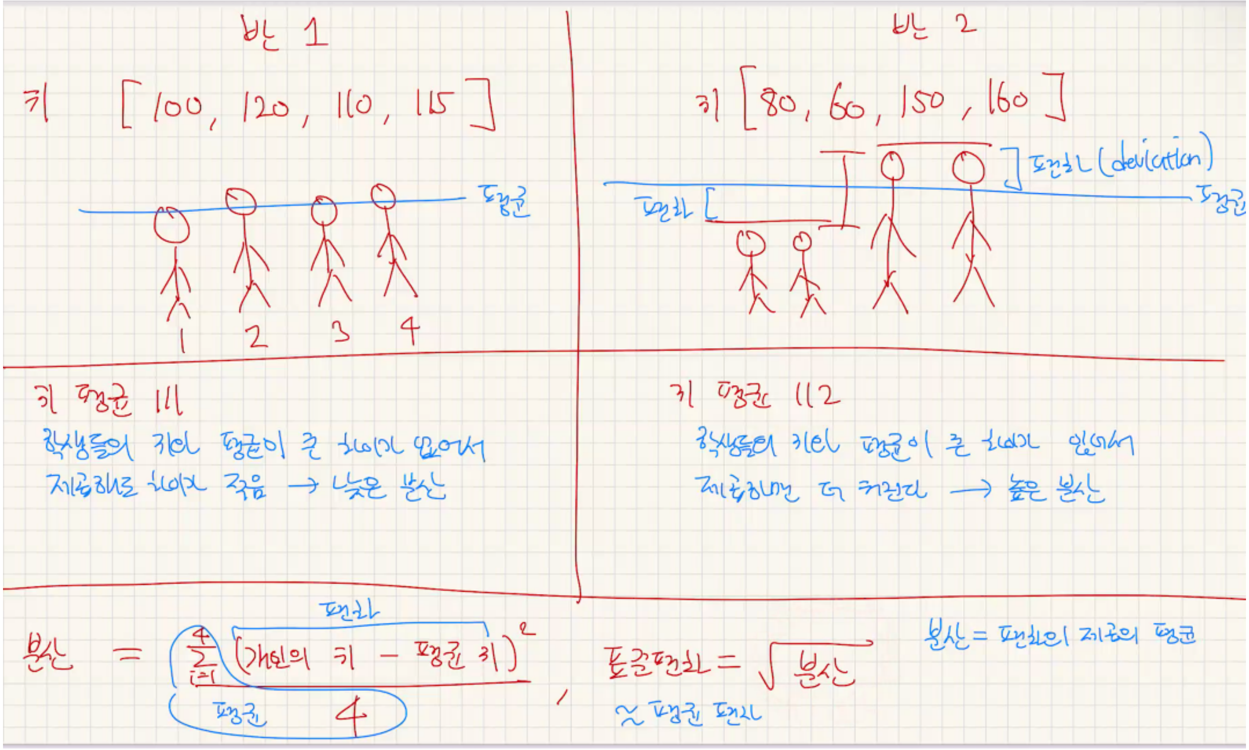
•상관계수는 공분산보다 해석이 용이합니다.

$$\boxed{\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}}$$

Y의 표준편차

X의 표준편차

분산과 표준편차(평균편차)



공분산과 상관계수의 차이점

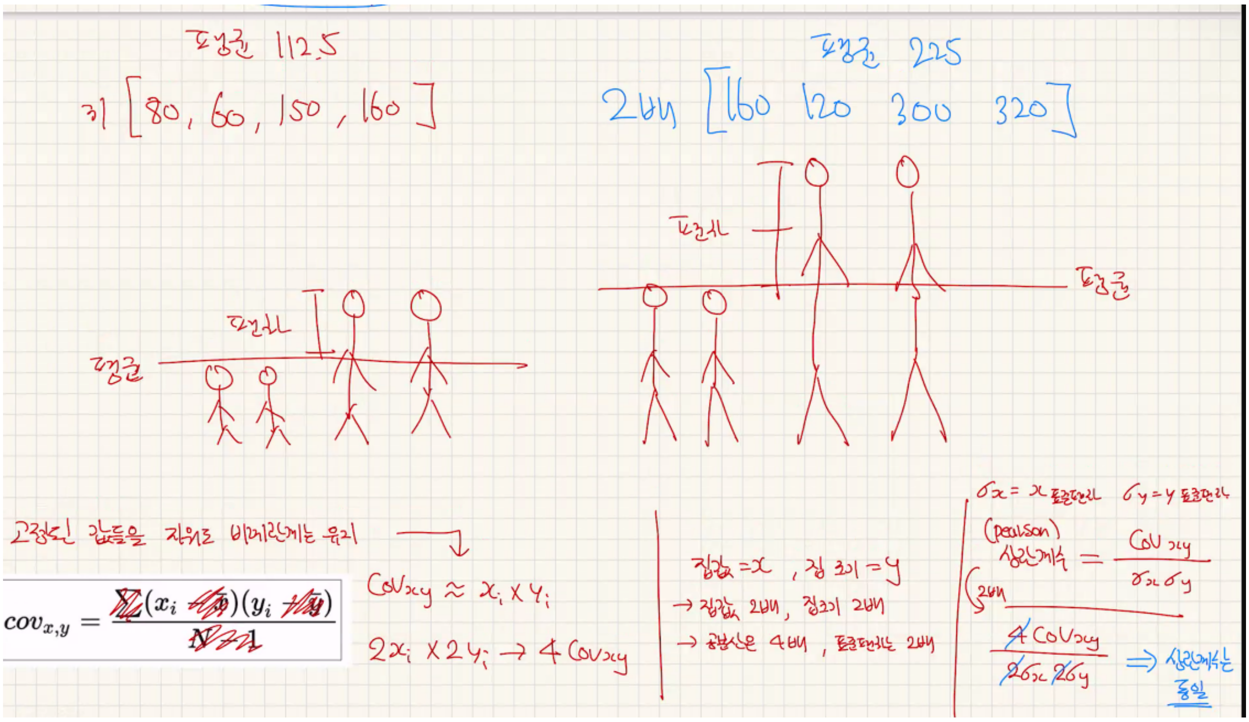
확률변수 x, y에 대하여

공분산: 근본적인 비례 관계가 동일하더라도, 단위가 달라지면 값도 달라진다 (단점)

상관계수: 단위가 달라도 변하지않는다.(단점극복)- 근본적인 비례 관계만 고려한다.

================================

예시



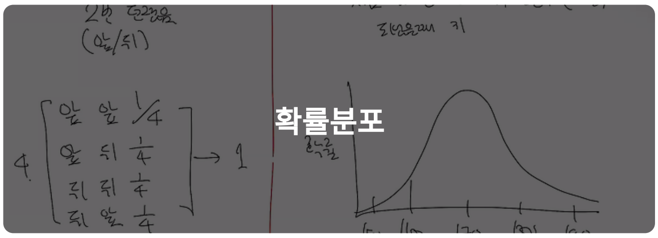
♡ 1 구독하기

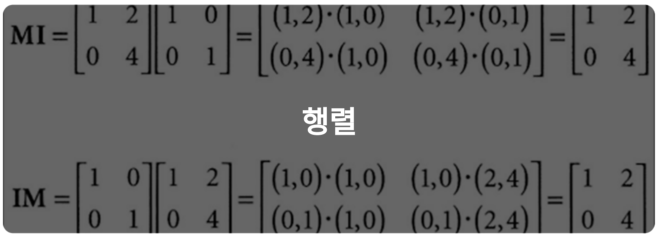
'practice_선형대수' 카테고리의 다른 글

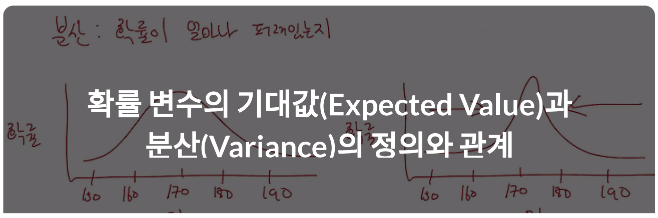
확률분포 (0)	2024.09.30
행렬 (1)	2024.09.30
확률 변수의 기대값(Expected Value)과 분산(Variance)의 정의와 관계 (0)	2024.09.29
최소 제곱법 (0)	2024.09.29
벡터공간과 성질 (0)	2024.09.28

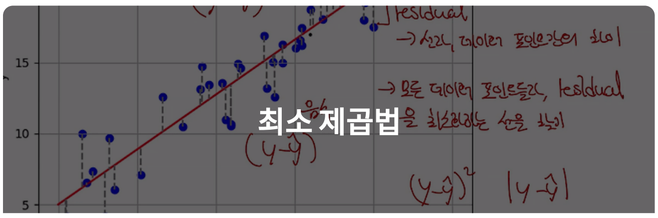
관련글

관련글 더보기


확률분포


행렬


확률 변수의 기대값(Expected Value)과 분산(Variance)의 정의와 관계


최소 제곱법

자연어(NLP)

네이쳐2024 님의 블로그입니다.

구독하기+

댓글 0

이름

비밀번호

내용을 입력하세요.

등록