IST108
OLASILIK VE İSTATİSTİK

KOVARYANS VE KORELASYON

İçerik

Kovaryans

Kovaryans, iki rastgele değişken arasındaki ilişkinin bir ifade biçimidir.

X ve Y iki rastgele değişken olsun.

μ_x ve μ_y sırasıyla X ve Y rastgele değişkenlerinin ortalamaları olsun.

X ve Y rastgele değişkenlerinin kovaryansı aşağıdaki şekilde tanımlanır.

$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$

Eşitliğin sağ tarafının açılımını yaparsak aşağıdaki ifadeyi elde ederiz.

$Cov(X,Y) = E[XY] - E[X]E[Y]$

Kovaryansın Özellikleri

$Cov(X,Y) = Cov(Y,X)$

$Cov(X,X) = Var(X)$

$Cov(aX,Y) = aCov(X,Y)$

$Cov(X+Z,Y) = Cov(X,Y) + Cov(Z,Y)$

$Cov(\sum_{i=1}^{n} X_i, Y) = \sum_{i=1}^{n} Cov(X_i, Y)$

Kovaryans ve Bağımlı Rastgele Değişken

Bağımlı rastgele değişkenlerin varyanslarının toplamı, kovaryansları kullanılarak aşağıdaki şekilde ifade edilebilir.

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} Cov(X_i, X_j)$$

İki rastgele değişken için

$$Var(X+Y) = Var(X) + Var(Y) + Cov(X,Y) + Cov(Y,X)$$

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

Bağımsız Rastgele Değişkenlerin Kovaryansı

X ve Y bağımsız rastgele değişkenlerse kovaryansları 0'a eşit olur.

$Cov(X,Y) = 0$

Neden?

X ve Y bağımsız rastgele değişkenlerse

$E[XY] = E[X]E[Y]$

Dolayısıyla

$Cov(X,Y) = E[XY] - E[X]E[Y]$

$Cov(X,Y) = E[XY] - E[XY] = 0$

Bağımsız Rastgele Değişkenlerin Kovaryansı

X ve Y bağımsız rastgele değişkenlerse kovaryansları 0'a eşittir. Ancak bunun tersi her zaman doğru değildir.

İki rastgele değişkenin kovaryansı 0'a eşit ise bu iki rastgele değişken bağımsızdır diyemeyiz.

Bağımlı rastgele değişkenlerin de kovaryansları 0'a eşit olabilir.

Kovaryans Ne İfade Eder?

$Cov(X,Y)$'nin değeri pozitif ise

- X arttıkça Y'nin artma eğiliminde olduğunun göstergesi

$Cov(X,Y)$'nin değeri negatif ise

- X arttıkça Y'nin azalma eğiliminde olduğunun göstergesi

Korelasyon

X ve Y iki rastgele değişken olsun.

Bu iki rastgele değişken arasındaki ilişkinin gücü korelasyon ile ifade edilebilir.

Korelasyon aşağıdaki şekilde tanımlanır.

$Corr(X,Y) = \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

Korelasyon her zaman -1 ile $+1$ arasında değer alır.

$-1 \leq Corr(X,Y) \leq 1$

Korelasyonun Özellikleri

$Corr(X,Y) = 1 \Rightarrow Y = aX + b\ ve\ a > 0$

$Corr(X,Y) = -1 \Rightarrow Y = aX + b\ ve\ a < 0$

$Corr(X,Y)$'nin değeri pozitif ise

- X arttıkça Y'nin artma eğiliminde olduğunun göstergesi

$Corr(X,Y)$'nin değeri negatif ise

- X arttıkça Y'nin azalma eğiliminde olduğunun göstergesi

$Corr(X,Y) = 0$ ise X ve Y rastgele değişkenleri **birbiri ile ilişkisizdir** denir.

Korelasyonun Özellikleri

Korelasyon, iki rastgele değişkenin arasındaki doğrusal ilişkiye dair bilgi verir.

- $Corr(X,Y) = 1$ ise veya $Corr(X,Y) = -1$ ise X ve Y rastgele değişkenleri arasında tam doğrusal ilişki var demektir.
- $Corr(X,Y)$, 1 veya -1 civarında ise örneğin 0,8 veya $-0{,}7$ gibi değer ise X ve Y rastgele değişkenleri arasında göreli olarak güçlü bir doğrusal ilişki var demektir.
- $Corr(X,Y)$, 0 civarında ise örneğin 0,3 gibi değer ise X ve Y rastgele değişkenleri arasında göreli olarak zayıf bir doğrusal ilişki var demektir.

Örnek 1

W, X, Y ve Z rastgele değişkenleri ikili olarak bağımsızdır.

$E[W] = E[X] = E[Y] = E[Z] = 0$ ve

$Var(W) = Var(X) = Var(Y) = Var(Z) = 1$ ve

$R = W + X, S = X + Y$ ve $T = Y + Z$ ise

A) $Corr(R,S) = ?$

B) $Corr(R,T) = ?$

Örnek 1

A) $Corr(R,S) = ?$

$Corr(R,S) = \frac{Cov(R,S)}{\sqrt{Var(R)Var(S)}}$

$Cov(R,S) = E[RS] - E[R]E[S]$

$= E[(W+X)(X+Y)] - E[W+X]E[X+Y]$

$= E[WX + WY + X^2 + XY] - (E[W] + E[X])(E[X] + E[Y])$

$= E[X^2]$

$= Var(X) + E[X]^2$

$= 1$

Örnek 1

A) $Corr(R,S) = ?$

$Corr(R,S) = \frac{Cov(R,S)}{\sqrt{Var(R)Var(S)}}$

$Cov(R,S) = 1$

$Var(R) = Var(W) + Var(X) = 2$

$Var(S) = Var(X) + Var(Y) = 2$

$Corr(R,S) = \frac{Cov(R,S)}{\sqrt{Var(R)Var(S)}} = \frac{1}{\sqrt{2\times 2}} = \frac{1}{2}$

Örnek 1

B) $Corr(R,T) = ?$

$Corr(R,T) = \frac{Cov(R,T)}{\sqrt{Var(R)Var(T)}}$

$Cov(R,T) = E[RT] - E[R]E[T]$

$= E[(W+X)(Y+Z)] - E[W+X]E[Y+Z]$

$= E[WY + WZ + XY + XZ] - (E[W] + E[X])(E[Y] + E[Z])$

$= 0$

$Corr(R,T) = \frac{Cov(R,T)}{\sqrt{Var(R)Var(T)}} = \frac{0}{\sqrt{Var(R)Var(T)}} = 0$

Örnek 2

X ve Y birer rastgele değişkendir.

$Y = aX + b$ ise, X ve Y arasındaki kovaryans ve korelasyonu hesaplayınız.

Örnek 2

$$\begin{aligned}Cov(X,Y) &= E[XY]-E[X]E[Y]\\ &= E[X(aX+b)]-E[X](aE[X]+b)\\ &= E[aX^2+bX]-aE[X]^2-bE[X]\\ &= aE[X^2]+bE[X]-aE[X]^2-bE[X]\\ &= aE[X^2]-aE[X]^2\\ &= aVar(X)\end{aligned}$$

Örnek 2

$Cov(X,Y) = aVar(X)$

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}=\frac{aVar(X)}{\sqrt{Var(X)a^2Var(X)}}=\frac{a}{|a|}$$

$$Corr(X,Y)=\begin{cases}1 & a>0\\ -1 & a<0\end{cases}$$

Örnek 3

X, rastgele değişkeni $(-1,1)$ arasında düzgün olarak dağıtılmıştır.

$Y = X^n$ ise X ve Y arasındaki kovaryans ve korelasyonu hesaplayınız.

Örnek 3

$$\begin{aligned}Cov(X,Y) &= E[XY]-E[X]E[Y]\\ &= E[XX^n]-\underbrace{E[X]}_{=0}E[X^n]\\ &= E[X^{n+1}]=\int_{-1}^{1}\frac{x^{n+1}}{2}dx=\frac{x^{n+2}}{2(n+2)}\Big|_{-1}^{1}=\frac{(1)^{n+2}-(-1)^{n+2}}{2(n+2)}\\ &= \begin{cases}\frac{1}{n+2} & n\ tek\ ise\\ 0 & n\ çift\ ise\end{cases}\end{aligned}$$

Örnek 3

$$Cov(X,Y) = \begin{cases} \frac{1}{n+2} & n\ tek\ ise \\ 0 & n\ çift\ ise \end{cases}$$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

n çift ise $Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{0}{\sqrt{Var(X)\,Var(Y)}} = 0$

n tek ise $Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{\frac{1}{n+2}}{\sqrt{\left(\frac{1}{3}\right)\left(\frac{1}{2n+1}\right)}} = \frac{\sqrt{3(2n+1)}}{n+2}$