상관계수

정의

- 공분산은 X와 Y가 함께 변해가는 기댓값을 말하고
- COV(x, y) = $E[(x-\mu_x)\,(y-\mu_y)]$로 정의된다. 즉

1) $E[(x-\mu_x)\,(y-\mu_y)]=E[xy - \mathrm{y}\mu_x - x\mu_y + \mu_x\mu_y]= E[xy] - \cancel{\mu_x \mathrm{E}}[\mathrm{y}] - \cancel{\mu_y E}[x] + E[\mu_x\mu_y]$

(1) 위 식을 정리하면, $E[(x-\mu_x)\,(y-\mu_y)]= E[xy] + \mu_x\mu_y$가 된다.

2) 상관계수는 공분산을 양 확률변수의 표준편차로 표준화한 값이다.

(1) $\dfrac{\mathrm{COV}(x,y)}{\sigma_x\sigma_y} = \dfrac{E[xy] + \mu_x\mu_y}{\sigma_x\sigma_y} = \rho$

정의

(2) 이를 이용하여 다음의 관계를 도출할 수 있다.

- $E[xy] = \rho\sigma_x\sigma_y + \mu_x \mu_y = E[x]E[y] + \mathrm{Cov}(\mathrm{x},\mathrm{y})$
- 즉, 확률변수의 곱의 기댓값 $E[xy]$ 는 공분산에 $E[x]E[y]$를 더한것과 같다.
- 또한, 공분산은 $\rho\sigma_x\sigma_y$와 같고,
- 또한, 공분산은 $E[xy]$ - $E[x]E[y]$와 같다.

정의

- 상관계수와 선형방정식의 계수

1. E $(y|x) = \int_{-\infty}^{\infty} y f_{y|x}(y|x)\, dy = \int_{-\infty}^{\infty} y \frac{f_{x,y}(x,y)}{f_x(x)} dy = \frac{1}{f_x(x)} \int_{-\infty}^{\infty} y f_{x,y}(x, y)\, dy$ 에서

1) 이를 선형방정식으로 본다면

(1) $\frac{1}{f_x(x)} \int_{-\infty}^{\infty} y f_{x,y}(x, y)\, dy$ = ax + b 에서 $\int_{-\infty}^{\infty} y f_{x,y}(x, y)\, dy$ = $f_x(x)$ (ax + b)

(2) 양변을 dx로 적분하면 $\int \int_{-\infty}^{\infty} y f_{x,y}(x, y)\, dxdy$ = $\int f_x(x)$ (ax + b)dx

- 이는 $\int \int_{-\infty}^{\infty} y f_{x,y}(x, y)\, dxdy = \int^{\infty} y f_y(y) dy$ = $\mathrm{E}(y)$ 이고

- $\int f_x(x)$ (ax + b)dx = E(ax+b) = aE(x) + b이다.

(3) 다시 정리하면, $\mathrm{E}(y) = \mathrm{aE}(x) + b$ 에서 $\mu_y = a\mu_x + b$

정의

2) 다시, 이번엔 $\int\int_{-\infty}^{\infty} y f_{x,y}(x,y)\, dxdy = \int f_x(x)$ (ax + b)dx 양변에 x를 곱하고 적분하면

(1) $\int\int_{-\infty}^{\infty} xy f_{x,y}(x,y)\, dxdy = \int x f_x(x)$ (ax + b)dx

(2) $\mathrm{E}(xy)= \int x f_x(x)$ (ax + b)dx $= \mathrm{E}(ax + bx^2)=$ $\mathrm{aE}(x) + bE(x^2)$

(3) 이 때, $\mathrm{E}(xy)= \rho\sigma_x\sigma_y + \mu_x\ \mu_y$ 이고, 따라서

- $\rho\sigma_x\sigma_y + \mu_x\ \mu_y$ = a$\mu_x + b({\mu_x}^2 + \sigma^2\ )$
- $a = \mu_y + \rho\ \frac{\sigma_y}{\sigma_x}\ \mu_x$
- b = $\rho\ \frac{\sigma_y}{\sigma_x}$

(4) 따라서 E $(y|x)= \mu_y\ + \rho\ \frac{\sigma_y}{\sigma_x}(\mathrm{x} - \mu_x)$ 이다.

정의

- 상관계수는 mgf를 통해 도출할 수 있다.

1) $\frac{M(t_1,t_2)}{\partial {t_1}^k \partial {t_2}^k} = \int\int_{-\infty}^{\infty} x^k y^k f_{x,y}(x,y)\,dxdy$ 에서

(1) E(x) = $\int\int_{-\infty}^{\infty} x f_{x,y}(x,y)\,dxdy$

(2) E(y) = $\int\int_{-\infty}^{\infty} y f_{x,y}(x,y)\,dxdy$

(3) E(x^2) = $\int\int_{-\infty}^{\infty} x^2 f_{x,y}(x,y)\,dxdy$

(4) E(y^2) = $\int\int_{-\infty}^{\infty} y^2 f_{x,y}(x,y)\,dxdy$

(5) $E[xy]$ = $\frac{M(t_1,t_2)}{\partial t_1 \partial t_1}$ = $\int\int_{-\infty}^{\infty} xy f_{x,y}(x,y)\,dxdy$

2) 따라서, $E[(x-\mu_x)\left(y-\mu_y\right)]$ = $E[xy] + \mu_x\mu_y$에서

(1) $\int\int_{-\infty}^{\infty} xy f_{x,y}(x,y)\,dxdy$+ $[\int\int_{-\infty}^{\infty} x f_{x,y}(x,y)\,dxdy]$·$[\int\int_{-\infty}^{\infty} y f_{x,y}(x,y)\,dxdy]$