

Chapter03

통계분석 I

I 상관성 분석

M T W T F S S

FASTCAMPUS
ONLINE

금융공학/퀀트 I

강사. 장순용

I 키포인트

- 피어슨, 켄달, 스피어맨 상관계수.
- 피어슨 상관계수의 신뢰구간.

I 피어슨 상관계수 (Pearson's correlation coefficient)

- “일상적인 상관계수”이고 다음과 같은 수식으로 계산할 수 있다.

$$Cor(X, Y) = \frac{Cov(X, Y)}{s_X s_Y}$$

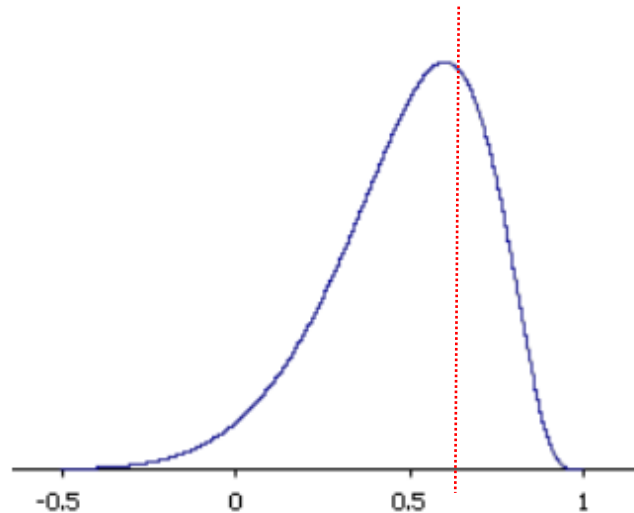
- 피어슨 상관계수의 값은 -1과 1사이의 수치이다.

I 피어슨 상관계수 (Pearson's correlation coefficient)

- 피어슨 상관계수는 선형관계의 방향과 강도를 나타낸다.
 - $\Rightarrow Cor(X, Y) > 0$: X 와 Y 사이에 **양**의 선형관계가 있음.
 - $\Rightarrow Cor(X, Y) < 0$: X 와 Y 사이에 **음**의 선형관계가 있음.
 - $\Rightarrow Cor(X, Y) = 0$: X 와 Y 사이에 선형관계가 **없음**.

I 피어슨 상관계수 (Pearson's correlation coefficient)

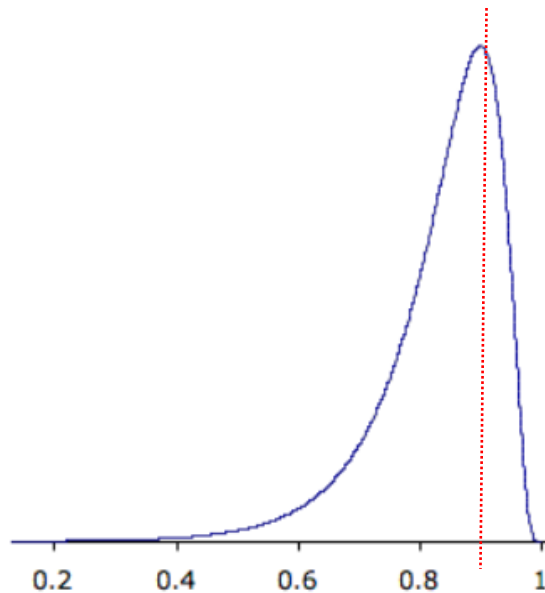
- 그런데 표본 상관계수 r 은 정규분포를 정확하게 따르지 않는다.



$$n = 12, \text{ 모상관계수} = 0.6$$

I 피어슨 상관계수 (Pearson's correlation coefficient)

- 그런데 표본 상관계수 r 은 정규분포를 정확하게 따르지 않는다.



$$n = 12, \text{ 모상관계수} = 0.9$$

I 피어슨 상관계수 (Pearson's correlation coefficient)

- 그런데 표본 상관계수 r 은 정규분포를 정확하게 따르지 않는다.
- 다음과 같이 변환된 수치는 정규분포를 따른다: “**피셔의 z 변환**”

$$\Rightarrow z = 0.5 \ln \left(\frac{1+r}{1-r} \right) = \operatorname{arctanh}(r)$$

$$\Rightarrow \sigma_z = \frac{1}{\sqrt{n-3}} \quad \Leftarrow \text{“표준오차”}$$

- 역변환:

$$\Rightarrow r = \frac{e^{2z}-1}{e^{2z}+1} = \tanh(z)$$

I 피어슨 상관계수 (Pearson's correlation coefficient)

- 다음과 같이 임의의 신뢰수준 확률 $1 - \alpha$ 에 해당하는 피어슨 상관 계수의 신뢰구간을 만들 수 있다:

하 한 : $\tanh\left(z - qnorm(1 - \frac{\alpha}{2}) \times \sigma_z\right)$

상 한 :

$\tanh\left(z + qnorm(1 - \frac{\alpha}{2}) \times \sigma_z\right)$

[

←

<신뢰구간>

→

]

I 스피어맨 상관계수 (Spearman's correlation coefficient)

- X 와 Y 변수의 순위(*rank*) 사이의 상관성을 나타낸다:

$$r_s = \frac{Cov(X_r, Y_r)}{s_{X_r} s_{Y_r}}$$

- 데이터에 순위만 정할 수 있다면 수치형 변수가 아니어도 적용 가능.
- 스피어맨 상관계수의 값도 -1과 1사이의 수치이다.
- 스피어맨 상관계수는 단조로움 (monotonicity)의 관계를 표현한다.

I 켄달 순위 상관계수 (Kendall's rank correlation coefficient)

- (x, y) 형태로 이루어진 데이터가 있을때 i 번째와 j 번째를 비교한다.

⇒ **부합**: $x_i < x_j$ and $y_i < y_j$ 또는 $x_i > x_j$ and $y_i > y_j$

⇒ **비부합**: $x_i < x_j$ and $y_i > y_j$ 또는 $x_i > x_j$ and $y_i < y_j$

- 켄달 순위 상관계수 r_k 는 다음과 같이 구한다:

$$r_k = \frac{(\text{부합 짝의 수}) - (\text{비부합 짝의 수})}{\frac{1}{2}n(n-1)}$$

I 켄달 순위 상관계수 (Kendall's rank correlation coefficient)

- 켄달 순위 상관계수의 값도 -1 과 1 사이의 수치이다.

I 끝.

감사합니다.

