

## 상관계수 정의

- 숫자형-숫자형 변수의 관계를 파악할 때 : 산점도(그래프), 상관계수(수치)
- 숫자형-숫자형 변수간의 강도를 수치로 표현하는 방법
- 상관계수는 (인과성이 아닌) 연관성 만 확인가능하다.

(연관성 안에 필요한 조건이 만족될때 인과성이 생길 수 있음)

## 상관계수 분석

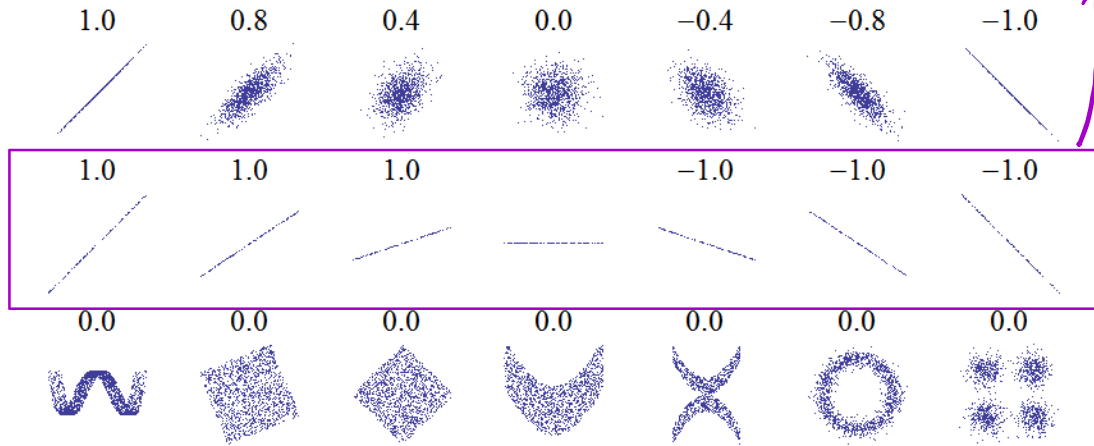
상관계수는 두 숫자형 변수사이의 연관성 중 직선적인 경향을 나타낸다.

즉, 직선을 띄느냐( -1 / +1 ) 아니면 퍼져있느냐( -1 ~ 1 ) 이다.

예를 들어, 기온이 45도 든 30도 든 60도 든 산점도상의 점들이 직선을 이룬다면 +1 아니면 -1이다. 즉, 직선형태를 이룬다면 상관관계는 1 혹은 -1(음의 기울기)이다.

직선적인 경향을 나타내는 척도이다.

- 직선이 아니라 흩어져있다면, -1 ~ 0, 0 ~ +1 값을 가지고, 가장 덜 직선같이 흩어져있으면 상관계수가 0이다.



위의 그림에서 상관값 = 0 은 가장 덜 직선 같이 생겼다. 강한 양/음의 상관 관계를 가진다면 -1 혹은 +1에 가깝다.

- 만약 세 변수의 상관관계를 파악하려고 한다면, 'multicorrelation' 공식을 이용하면 된다.

공식 내부를 보면, 나올 수 있는 모든 변수 짝의 상관관계를 분석한다.

co: 서로 간의  
relation: 관계

### ① Pearson correlation coefficient

수축기 혈압과 이완기 혈압, 키와 몸무게와 같이 분석하고자 하는 두 변수가 모두 연속형 자료일 때 두 변수간 선형적인 상관 관계의 크기를 모수적(parametric)인 방법으로 나타내는 값입니다.

“피어슨의 적률 상관 계수(Pearson's product moment correlation coefficient)”, 피어슨의  $r$ (Pearson's  $r$ )”, “ $r$ ”, “ $R$ ” 등은 모두 피어슨 상관 계수를 나타내는 다른 용어들입니다.

피어슨의 상관 계수의 크기는 절대적인 것은 아니지만 보통 아래와 같은 범위를 지정하여 두 변수간 상관 관계의 크고 작음을 나타냅니다.

-1.0  $\leq r \leq$  -0.7 : 매우 강한 음(-)의 상관 관계  
-0.7  $< r \leq$  -0.3 : 강한 음(-)의 상관 관계  
-0.3  $< r \leq$  -0.1 : 약한 음(-)의 상관 관계  
-0.1  $< r \leq$  0.1 : 상관 관계 없음  
0.1  $< r \leq$  0.3 : 약한 양(+)의 상관 관계  
0.3  $< r \leq$  0.7 : 강한 양(+)의 상관 관계  
0.7  $< r \leq$  1.0 : 매우 강한 양(+)의 상관 관계

### 상관 계수 검정

cor.test()를 사용하여 상관 계수 검정 Correlation Test을 수행하여 상관 계수의 통계적 유의성을 판단할 수 있다. 이때 귀무가설은 ‘ $H_0$ : 상관 계수가 0이다’이며, 대립가설은 ‘ $H_1$ : 상관 계수가 0이 아니다’이다.

$r=0$ 은 두 변수 사이에 직선적 상관관계가 없음을 의미하며, 0상관(zero correlation) 또는 무상관(no correlation)이라고 하는 한 가지 유의할 점은  $r=0$ 이 두 변수 간에 직선적 관계가 없음을 가리키는 것이지, 두 변수 사이에 어떤 관계도 존재하지 않는다는 뜻이 아니라는 것이다. 따라서 상관분석을 할 때는 먼저 산점도를 그려서 두 변수간의 관계를 미리 알아보는 것이 중요하다.

## 2단계: 상관 계수가 유의한지 여부 확인

두 변수 간의 '상관 계수'를 구했는데, 이것이 우연히 발생한 값인지 아닌지를 검증해야 함.  
변수 사이에 유의한 상관 관계가 있는지 여부를 확인하려면 p-값을 유의 수준과 비교하십시오. 일반적으로 0.05의 유의 수준( $\alpha$  또는 알파로 표시함)이 ~~중요한 기준~~ 0.05의  $\alpha$ 는 실제로 상관 관계가 존재하지 않는데 상관 관계가 존재한다는 결론을 내릴 위험이 5%라는 것을 나타냅니다. p-값을 통해 상관 계수가 0과 유의하게 다른지 알 수 있습니다. 상관 계수가 0이면 선형 관계가 없음을 나타냅니다.

**p-값  $\leq \alpha$ : 상관 관계가 통계적으로 유의합니다.**

p-값이 유의 수준보다 작거나 같으면 상관 계수가 0과 다르다는 결론을 내릴 수 있습니다.

**p-값  $> \alpha$ : 상관 관계가 통계적으로 유의하지 않습니다.**

p-값이 유의 수준보다 크면 상관 관계가 0과 다르다는 결론을 내릴 수 없습니다.

질문 중간에 답글을 쓰겠습니다.

>통계분석 책을 찾아보니까요..

>

>영어점수와 수학점수의 평균 차이를 예로 들면서

>상관계수가 .858로 높게 나왔는데 p값때문에 통계적으로 유의미하지 않은 결과가 나왔다.

>결과를 종합하면 통계적으로는 유의미하지는 않지만 영어와 수학점수 간의 상관관계가 높으면 이러한 관계는 통계적으로 유의미하다.

>따라서 영어를 잘하는 학생은 수학도 잘한다고 할 수 있다.

>

>이렇게 적혀있거든요?

Answer >

이런 경우가 가끔씩 나오게 됩니다. ① 원인은 여러 가지가 있을 수 있지만 가장 대표적인 이유는 data의 수가 작을 때입니다.

② 또 한가지는 이상값이 존재할 경우입니다.

위와 같은 경우에는 상관계수는 상당히 높는데 p 값은 0.05보다 큰 서로 모순적인 모습을 보이게 됩니다.

그러므로, 이런 경우에는 data의 수를 늘리거나 이상값을 찾아서 제거를 해 준 다음 다시 분석을 하면 됩니다. 거의 80% 정도는 여기에 해당된다고 생각할 수 있습니다.

>그럼 상관계수가 결과해석에 중요한 요인이 된다는 얘기죠?

>

>제가 분석한 자료는 하나는 상관계수가 .742로 높은 반면 p값 때문에 유의미하지 않은 결과가 있구요..

>

>다른 하나는 상관계수가 -.039로 별로 높지 않은 반면 p값이 .005로 매우 유의미한 결과(이것도 이렇게 해석하는 것이 맞는지 잘 모르겠어요)가 나왔거든요?

Answer >

이 경우도 마찬가지로 생각할 수 있습니다. 상관계수는 아주 작는데 p 값은 0.05보다 작아서 유의한 영향을 주는 경우는 data의 수가 너무 많거나, 이상값이 있을 경우에 자주 나오는 현상입니다.

## < ADSP 기준표 >

상관계수 범위	해 석
$0.7 < \gamma \leq 1$	강한 양(+)의 상관이 있다
$0.3 < \gamma \leq 0.7$	약한 양(+)의 상관이 있다
$0 < \gamma \leq 0.3$	거의 상관이 없다
$\gamma = 0$	상관관계(선형, 직선)가 존재하지 않는다
$-0.3 \leq \gamma < 0$	거의 상관이 없다
$-0.7 \leq \gamma < -0.3$	약한 음(-)의 상관이 있다
$-1 \leq \gamma < -0.7$	강한 음(-)의 상관이 있다



