

[10] 상관분석

1. 상관계수

- 1) 상관계수 : 두 변수 사이의 관계를 반영하는 수치 (-1에서 1 사이의 값)
- 2) 상관관계 : 두 변수 간의 관계

- 방향과 크기를 알 수 있음 (따라서, 공분산을 쓰는 것보다 많은 정보를 알 수 있음)
- ++) 공분산 : +/-로 방향은 알 수 있지만, 크기는 알 수 없음
- 1에 가까울수록 상관계수의 강도가 크다 / 1이면 직선
- 0이면 상관관계가 없음

3) 상관계수를 해석할 때 주의할 점

- 양수와 음수 상관없이 절대값으로 판단

4) 피어슨 상관계수 r

- rxy는 변수 x와 변수 y간의 상관관계

5) 상관계수 계산하기

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \div \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1} \times \frac{\sum (y_i - \bar{y})^2}{n - 1}}$$

2. 그래서 분자가
-값도 나와야 한다.

1. 제공된 값으로,
항상 +값만 나온다.

- $r_{xy} = S_{xy} / \sqrt{S_{xx} * S_{yy}}$

- 엑셀 함수 : CORREL

- 상관계수의 시각적 표현 : 산점도

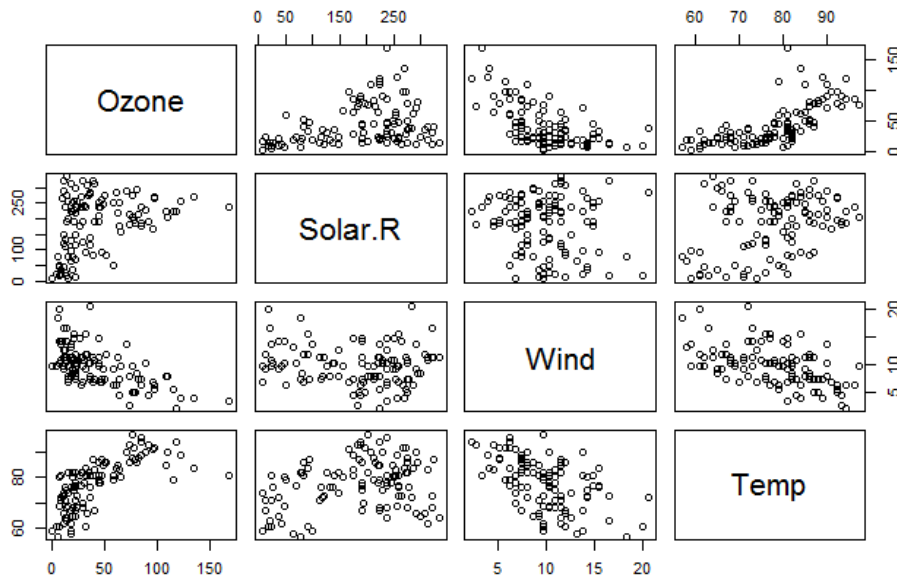
++) 산점도 엑셀로 : 삽입 > 차트 > 드롭다운 > 분산형 차트 유형

6) 상관계수 해석하기

상관계수 크기	일반적인 해석
±0.8 ~ 1.0	매우 강한 상관관계
±0.6 ~ 0.8	강한 상관관계
±0.4 ~ 0.6	중간 정도 상관관계
±0.2 ~ 0.4	약한 상관관계
±0.0 ~ 0.2	매우 약한 상관관계

7) 상관행렬

- 대각선의 값은 1
- 대각선을 중심으로 거울 같은 이미지를 만듦



**** 파이썬으로 해보기**
(코드 참고)

2. 결정계수

: 상관계수의 제곱

1) 다양한 상관계수

- 스피어만 서열 계수(스피어만 상관계수): 비모수통계분석
- 피어슨 상관계수: 모수통계분석

++) 모수, 비모수 통계분석 관련 : <https://kwon-coach.tistory.com/44>

****문제풀이**
(코드 참고)

3. 상관계수에 대한 t검정

1) 핵심 원리 알아보기

- 집단 간의 차이가 아니라 변수 간의 관계를 조사
- 두 개의 변수만 포함
- 적절한 검정통계는 상관계수에 대한 t검정

2) 검정통계량 계산하기

- r : 표본상관계수
- ρ : 모상관계수
- n : 표본의 크기
- $n-2$: 자유도(변수의 개수)

**** 설명할 수 있는 부분 : $1-r^2$**

$$t = \frac{r - \rho_0}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

3) 가설검정 과정

a. 귀무가설과 연구가설의 진술

- $H_0 : \rho_{xy} = 0$

- $H_1 : \rho_{xy} \neq 0$

b. 귀무가설과 관련된 위험 수준(또는 유의수준 또는 1종 오류 수준) 설정

c. 적절한 검정통계의 선택, 상관계수에 대한 t검정

d. 특정 통계에 대한 적절한 임계값 표를 사용하여 귀무가설을 기각하는 데 필요한 값 결정

- 자유도 : $n-2$

- 계산된 자유도($n-2$)와 감구하려는 위험 수준($\alpha=0.05$) 및 양측/단측 검정(연구가설에 방향이 없으면 양측 검정)을 사용하여 임계값 구하기

- 정리 : 따라서 위험 수준($\alpha=0.05$)에서 자유도가 $n-2$ 인 양측/단측 검정에 대해 귀무가설을 기각하는 데 필요한 값은 임계값이다~

e. 검정통계량(획득된 값)과 임계값의 비교

- 획득된 검정통계량 t 값은 ~이고 두 변수가 관련이 없다는 귀무가설을 기각하는 데 필요한 임계값은 ~이다.

f. 결정

** 문제 풀이

(엑셀/코드 참고)

- 엑셀에서 PEARSON 함수 사용해서 피어슨 상관계수도 구할 수 있음

- 결정계수는 상관계수 구해서 *2 하기

- t분포표에서 임계값 찾을 때, 양측검정 임계값은 $\alpha/2$ 줄에서 찾기

- 결과 ex. 임계값 > 검정통계량 이면 기각역이므로, 귀무가설을 기각하고 대립가설을 채택
>> 통계적으로 유의한 상관관계가 있다 / 상관계수가 통계적으로 유의하다 / 통계적으로 유의한 상관계수가 있다