집고 넘어가기

상관관계

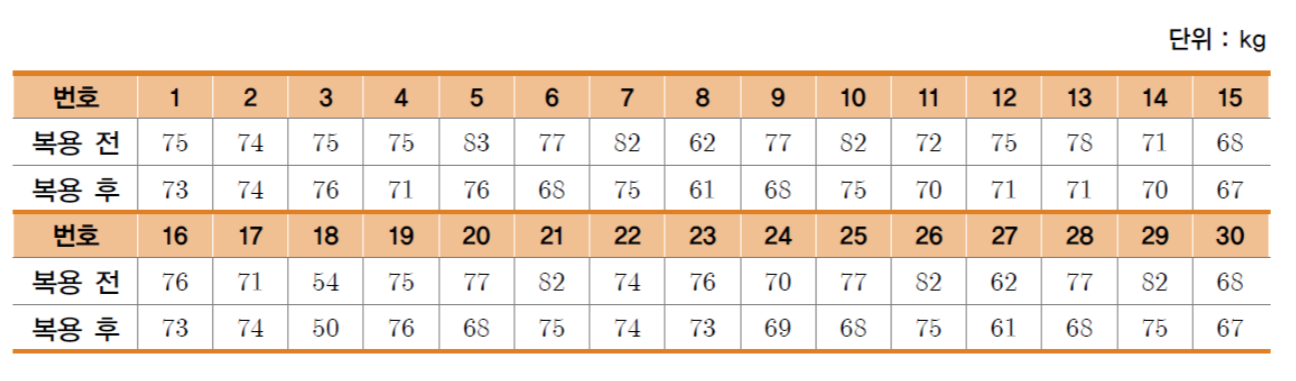
* 연관성을 확인할 수 있으며 숫자의 표현 가능 (-1 ~ 1)
* 원인과 결과형태인 인과관계가 아님

T검정

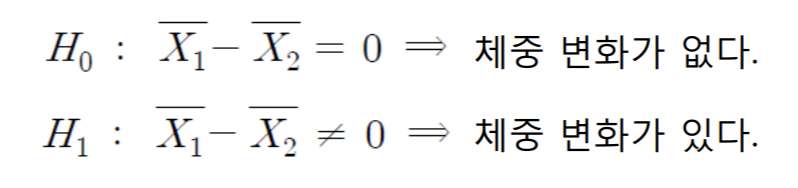
대응표본 T검정

* 두 모집단으로부터 표본을 추출할 때, 서로 연관된 표본

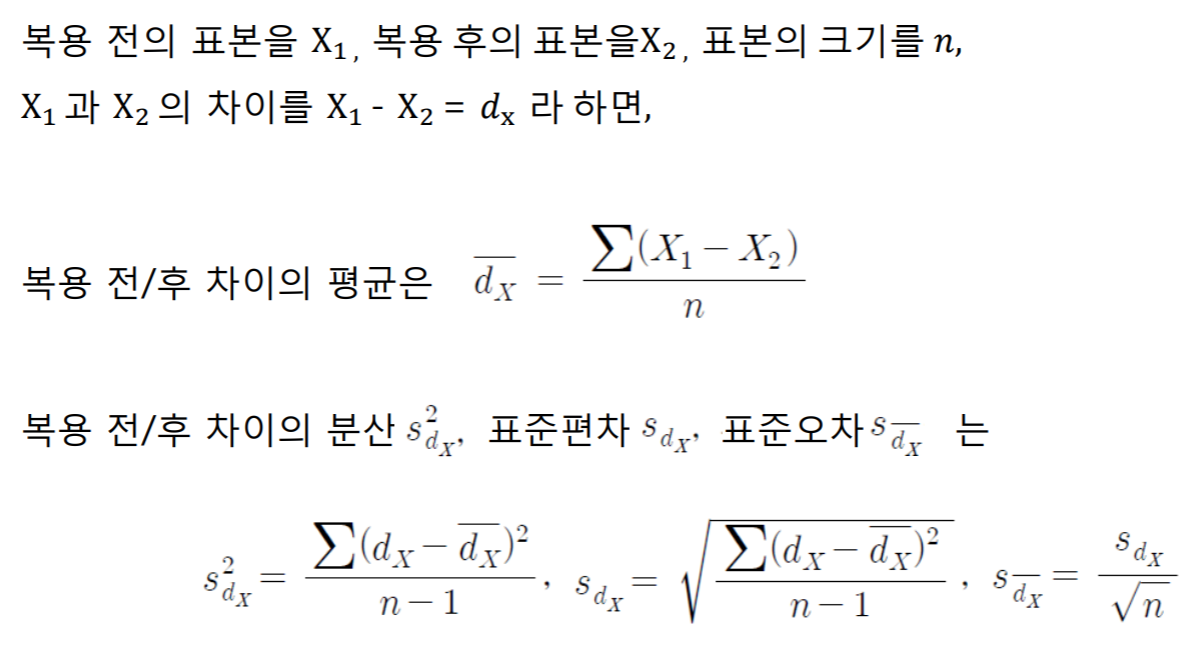
샘플데이터



가설



대응표본 검정 식

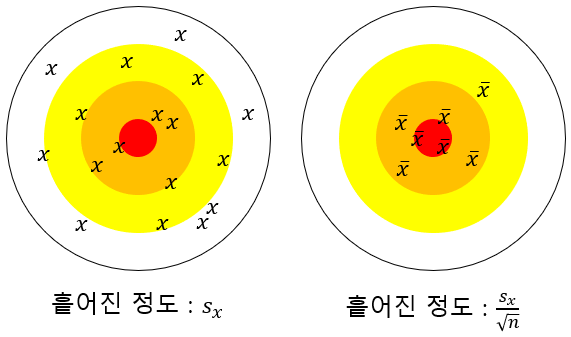


주의!

표준편차와 표준오차의 차이

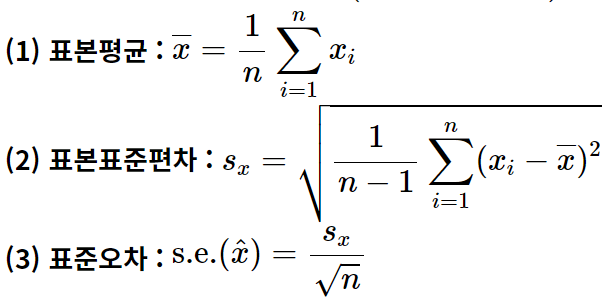
편차(변량-평균) : 제곱의 평균에 루트를 씌운 값이 표준편차

오차(추정값-참값) : 제곱의 평균에 루트를 씌운 것이 표준오차



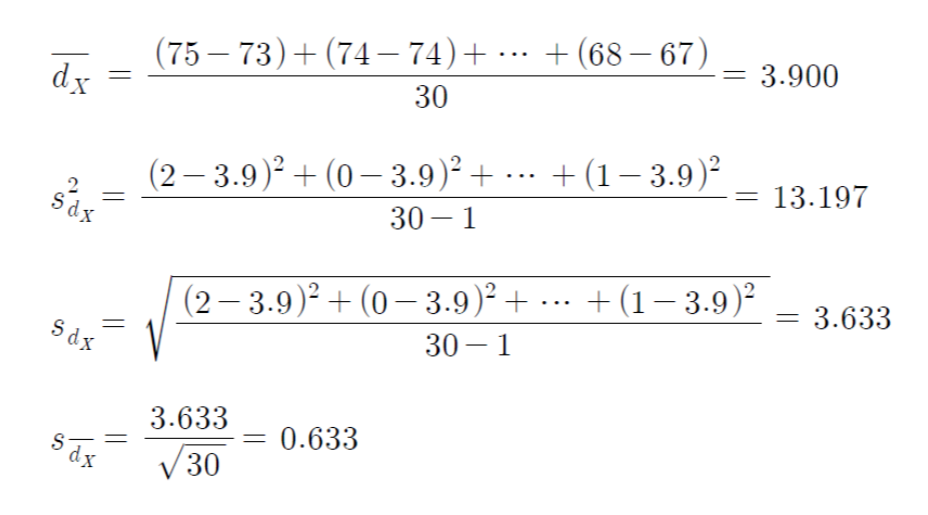
위의 그림처럼 표본을 계속해서 뽑게 되면 편차와 같이 표본평균으로부터 여러 표본들과의 오차가 생기게된다.

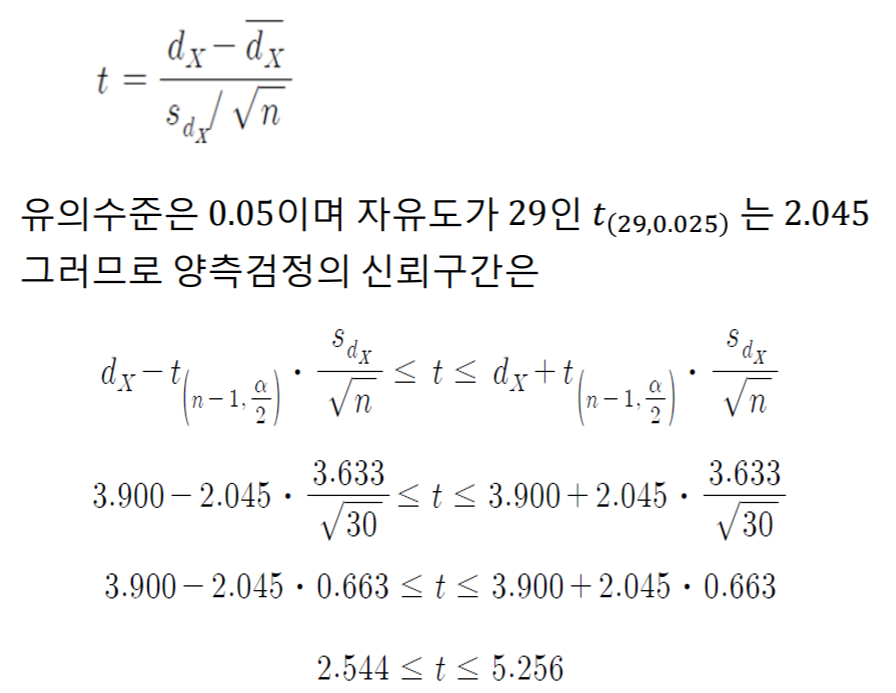
즉, 쉽게 얘기해서 표준편차를 계산할 때 루트 N으로 나눠주게 되면 표본평균의 표준편차가 된다.



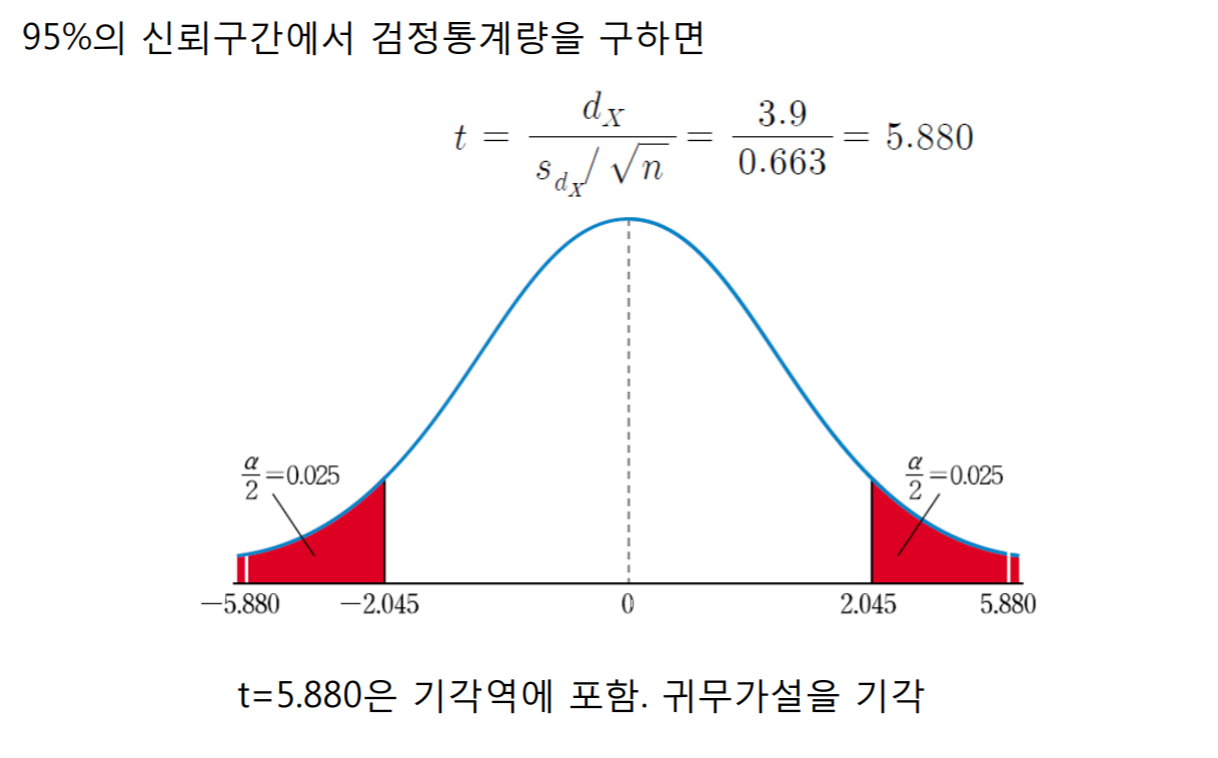
정의가 궁금하면 ? <https://hsm-edu.tistory.com/16>

다시 돌아와서 계산을 해보면





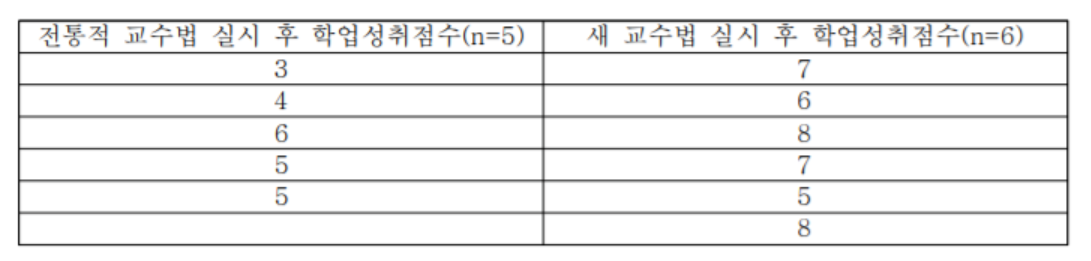
사설검정 결과는



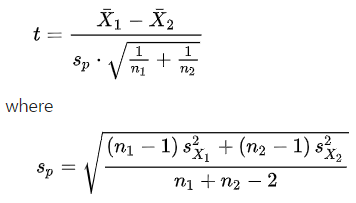
즉, T값은 5.88이고 신뢰구간 밖으로 나가니깐 귀무가설기각

사전 사후 차이가 있다는 것을 확인할 수 있다.

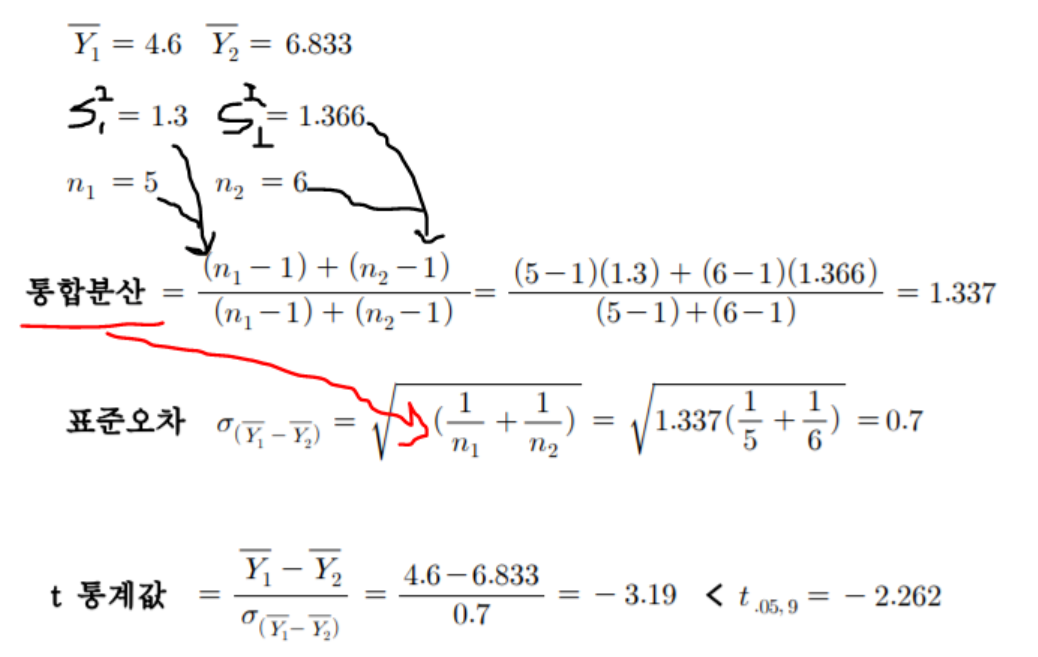
독립표본 T검정



독립 표본 T검정 식



위의 식을 기반으로 적용해본 결과 다음과 같다.



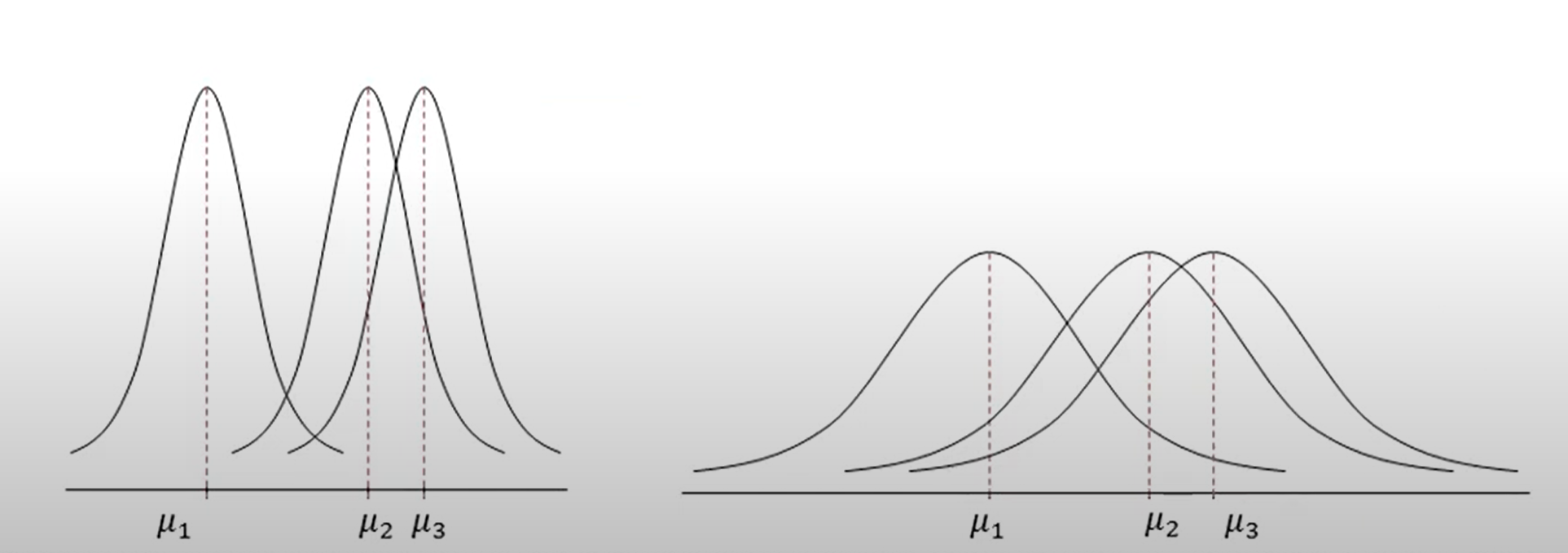
분산분석

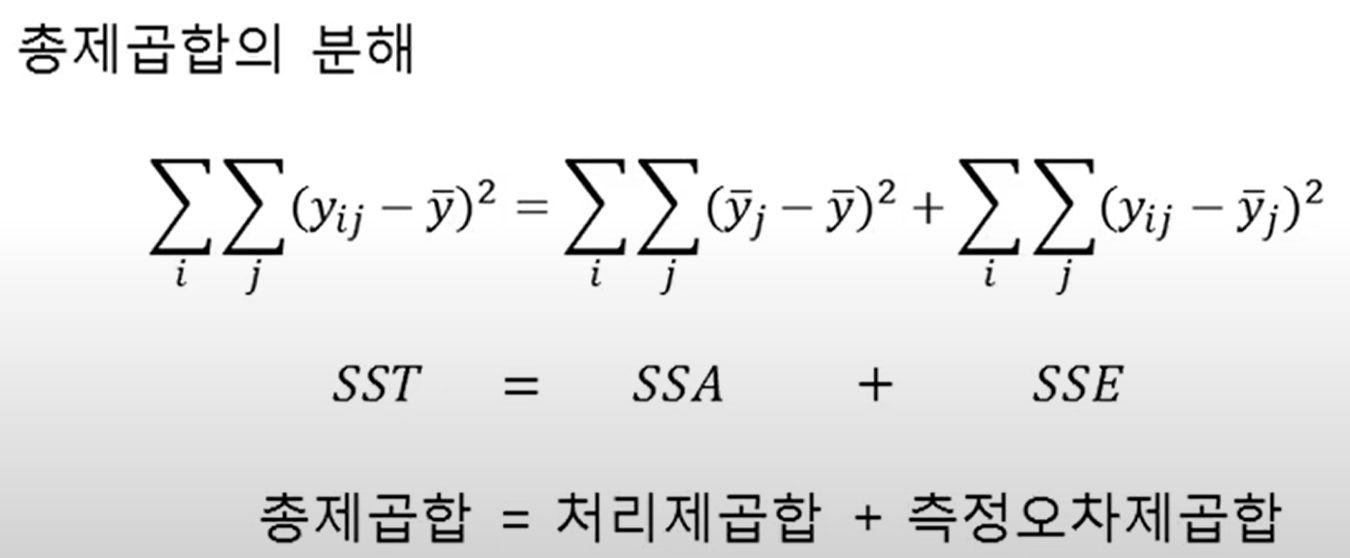
t검정을 여러 번 하면 되지 않을까??

* 여러 번 t-test를 해주면 1종오류(a)를 범할 확률이 증가
* 한번 가설검정을 할 경우마다 5% 유의수준이 중첨되기 때문에!

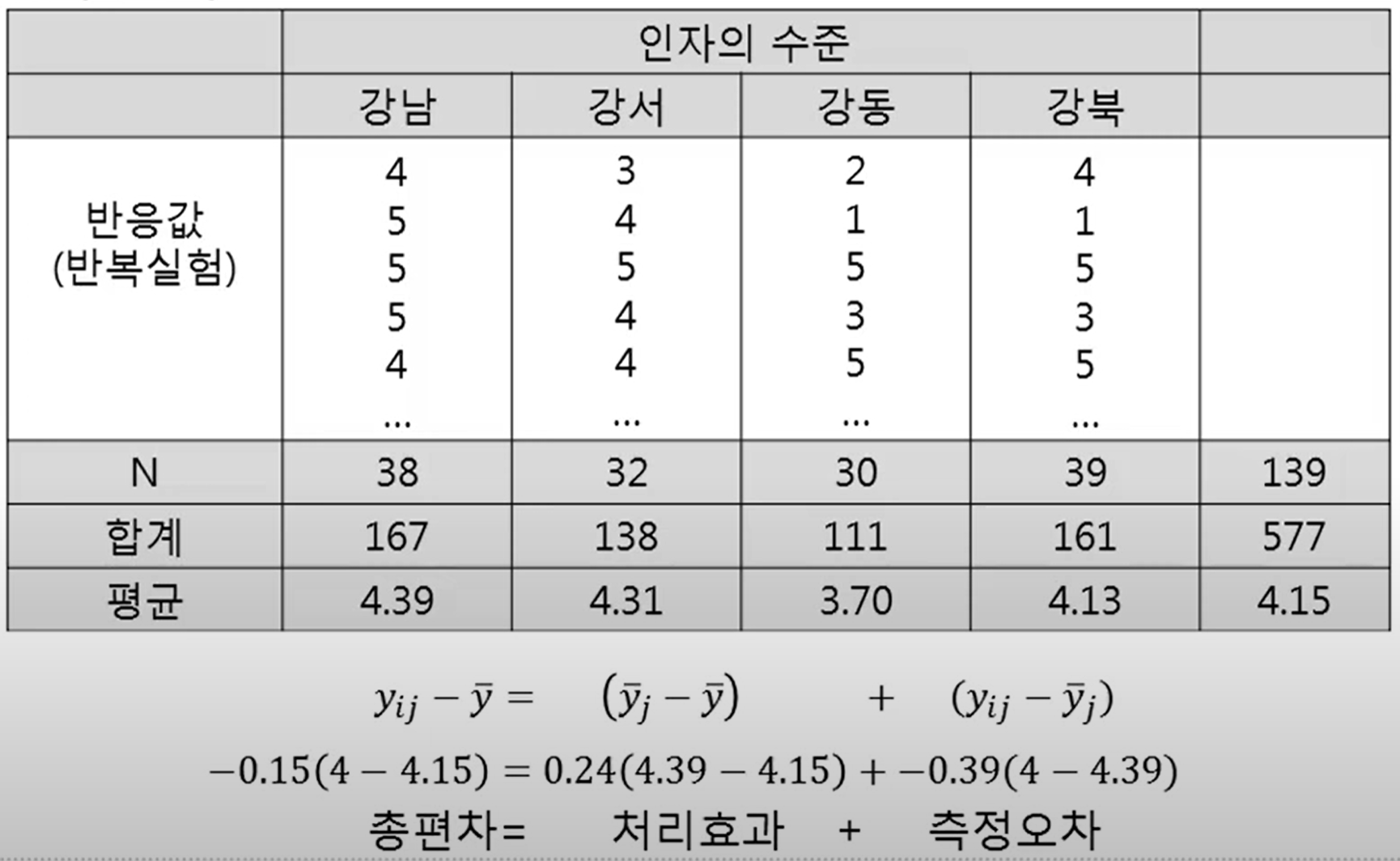
그룹내 분산의 비율을 통해 평균의 차이를 비교할 수 있다.

ANOCA(Analysis of Variance)





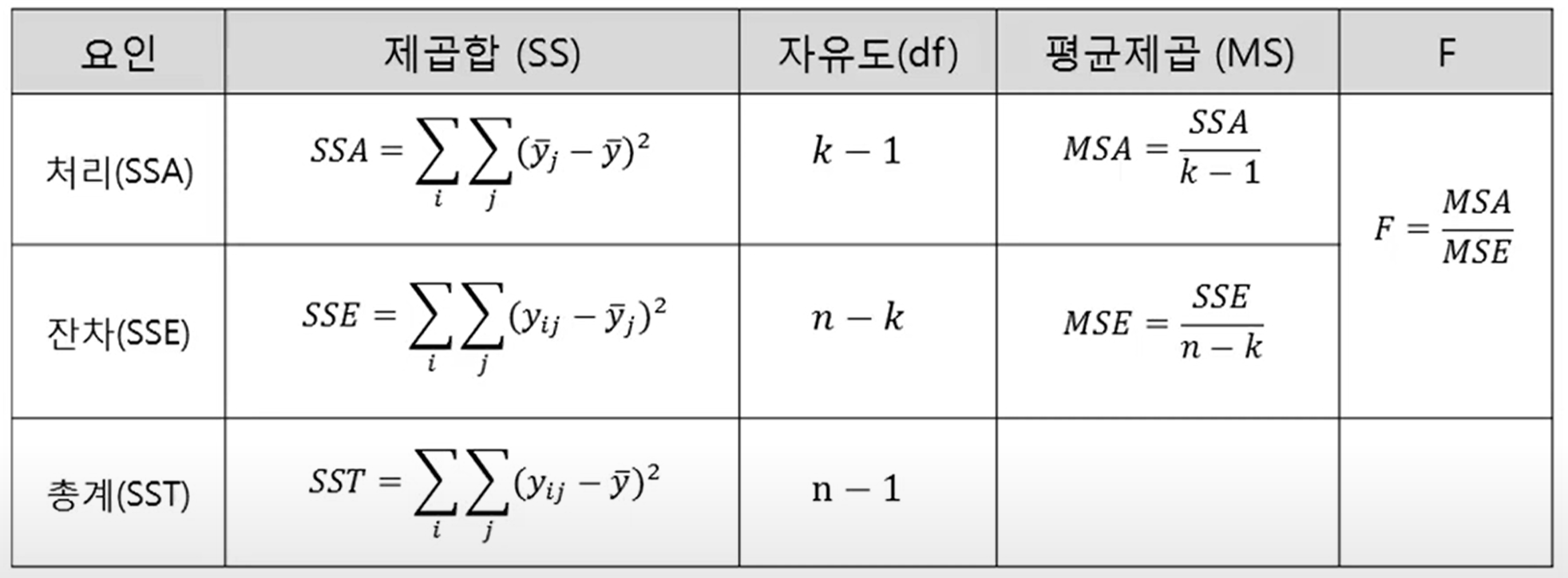
예시를 보면

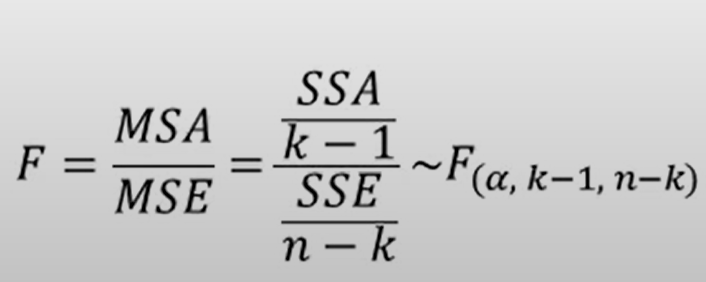


그룹간의 차이 = 처리효과부분

그룹내의 차이 = 측정오차부분

분산분석표





다음과 같이 F검정을 진행하는 것이다

그룹내의 분산과 그룹간의 분사을 통해서 평균을 기준으로 분산분석을 통해 검정을 하는 것이며

귀무가설 : 그룹간의 차이는 없다

대립가설 : 적어도 하나의 그룹간의 차이가 있다.

