

F검정 (분산분석, ANOVA)

F-검정값은 항상 1보다 같거나 크다. (두 표본집단의 분산 값을 나눈 것이므로, 큰 것이 분모, 작은 것이 분자)
F값이 클수록 두 집단간의 분산의 차이가 존재하는 것을 의미.

F검정(분산분석)의 이해

두 집단의 평균을 비교할 때는 t 검정을 쓰고, 3개 이상의 집단의 평균을 비교할 때는 분산 분석을 수행한다.

두 집단의 경우에 분산분석을 해도 되는데, 그때는 t 검정과 결과가 동일함

세 개 이상의 집단을 대상으로 두 집단씩 여러번 t검정 을 수행하게 되면 신뢰도가 급격히 감소하게 되는데 분산 분석은 이러한 현상이 발생하지 않는다.

분산

- 데이터의 변동 정도를 의미하는 지표
- 편차의 제곱에 대한 평균
- 분산이 크다 --> 데이터의 산포도가 크다 (데이터가 넓게 퍼져있음)

분산분석의 조건

조건	설명
정규성	각 샘플 데이터는 정규분포를 따르는 모집단으로 부터 추출되었다
등분산성	Y의 모집단 분산은 각각의 모집단에서 동일

조건	설명
독립성	그룹들의 모집단의 분산은 모두 동일하다

분산 분석의 종류

종류	이름	설명
일원분산 분석	One-way ANOVA	종속변수1개, 독립변수 집단도 1개인 경우 한가지 변수의 변화가 결과 변수에 미치는 영향을 알아보기 위해서 사용됨
이원분산 분석	Two-way ANOVA	독립변수 집단이 두 개 이상인 집단간 차이가 유의한지를 검증하기 위해 사용 한 변수의 변화가 결과에 미치는 영향이 다른 변수의 수준에 따라 달라지는지를 확인하기 위해 사용

F검정의 가설

가설	설명	식
귀무가설	모든 집단의 평균이 같다	$H_0 : \sigma_1 = \sigma_2$
대립가설	어떤 집단의 평균이 다르다	$H_0 : \sigma_1 \neq \sigma_2$

$p < \text{유의수준}$ 이면 귀무가설을 기각하고, 대립가설을 채택한다.

대립가설을 자세히 보시면 "어떤 집단의 평균이 다르다"이다.

집단이 A, B, C가 있으면 A와 B가 다른지 B와 C가 다른지는 분산 분석으로 알 수 없다.

A와 B가 다른지 B와 C가 다른지는 사후 검정으로 확인해야 한다.

파이썬에서 ANOVA를 사용하기 위한 패키지

`scipy.stats` 패키지와 `statsmodels` 패키지가 ANOVA 검정 방법을 제공한다.