

사회통계연습

ANOVA 연습

김현우, PhD¹

¹충북대학교 사회학과 부교수



진행 순서

- 1 t 검정과 일원분산분석
- 2 일원분산분석의 분석 사례
- 3 일원분산분석 연습



t 검정과 일원분산분석



t 검정과 일원분산분석

t 검정과 일원분산분석에는 잘 알려진 연관성이 있다.

- t 검정은 두 모집단의 **평균을 비교**하고, 일원분산분석은 여러 모집단에 걸친 **분산의 비율**을 비교한다.
- 만일 집단이 두 개만 주어졌을 때 일원분산분석을 수행하면 어떤 결과를 가져올까?
- 이 경우 일원분산분석의 귀무가설은 “모든 집단에 걸쳐 평균값이 동일하다”였으므로 이는 다시 “두 집단에 걸쳐 평균값이 동일하다”로 축소된다.
- 즉 t 검정과 같은 것이 된다. 실제로 F 값과 t 값에는 다음과 같은 관계가 있다.

$$\sqrt{F} = |t| \quad (\text{또는 } F = t^2)$$



t 검정과 일원분산분석

연습 1. neighbor.sav에서 내국인/외국인 여부에 따라 주민과의 갈등 수준에 차이가 있는지 검정하시오. 이때 t 검정과 일원분산분석을 각각 수행하고 결과가 어떻게 다른지 확인하시오.



t 검정과 일원분산분석

- 그렇다면 반대로 생각해서, 집단이 여러 개 있을 때 구태여 일원분산분석 대신 t 검정을 여러 번 하면 안될까?
- 결론만 말하자면 (1) 굉장히 불편하고 혼란스러울 뿐 아니라, (2) 추정상의 오류를 저지르게 될 위험이 극단적으로 커지므로 권할 수 없다.
- 먼저 t 검정을 아주 여러 번 수행하고 비교해야 하는 부담이 있다.
- 예컨대 겨우 5개의 모집단을 비교하기 위해서 t 검정을 10번이나 수행해야 한다 (Why?).
- 이것은 기하급수적으로 증가하여 6개의 모집단을 비교하기 위해서는 t 검정을 15 번이나 수행해야 한다(Why?).



t 검정과 일원분산분석

- 게다가 이 10번의 t 검정을 수행하는 과정에서 최소 1번 이상 오류가 나타날 가능성 또한 급격히 증가한다.
- 예컨대 5% 유의확률이라면 1회 이상의 오류 확률은 약 40%나 된다(Why?).
- **이항분포(binomial distribution)**를 통해 이 확률분포를 계산할 수 있다. 즉 발생확률 p 가 0.05인데 10번의 시행 중 사건이 전혀 발생하지 않을 확률분포는 다음과 같다.

$$1 - \binom{n}{k} p^k (1-p)^{n-k} = 1 - \binom{10}{0} \cdot 0.05^0 \cdot (1-0.05)^{10-0}$$

- 이 사실을 간단히 엑셀에서 계산해보자.



일원분산분석의 용도



일원분산분석의 용도

경험적 연구에서 일원분산분석은 크게 두 부분에서 주로 활용된다.

- 첫째로 표본에 관한 **기술통계(descriptive statistics)**를 제시할 때 사용될 수 있다.
- 기술통계로 제시한다면, 표본 안의 핵심이 되는 범주형 관심변수 내지 종속변수에 따라 다른 여러 변수들이 어떻게 다른지 일원분산분석을 통해 보여줄 수 있다.
- 이때 범주형 변수는 명목형 내지 순서형 척도로 측정된 것이며 예컨대 최종학력, 출신지역, 지지하는 정당 등을 생각해 볼 수 있다.
- 지금까지 연습한 일원분산분석은 주로 이에 관한 것이고, 실무 수준에서 가장 폭넓게 사용된다.



일원분산분석의 용도

- 두번째로 회귀분석(regression analysis)에서 모형 적합도(goodness-of-fit) 지표 중 하나로 사용될 수 있다.
- 이 경우 귀무가설은 “모든 회귀계수들이 0이다”로 만일 이 귀무가설을 기각하지 못한다면 모델에 포함된 어떠한 독립변수 X 로도 종속변수 Y 를 의미있게 설명하지 못함을 의미한다(Why?). 당연히 이 경우에는 회귀모형을 처음부터 다시 만들어야 한다.
- 회귀분석의 맥락에서 대립가설은 “적어도 하나 이상의 회귀계수는 0이 아니다” 임에 주의할 것.
- 우리는 나중에 회귀분석을 배우면서 일원분산분석이 회귀분석의 맥락에서도 다시 한번 쓰이게 됨을 확인하게 된다.



일원분산분석 연습



연습 2. FPOLICY.SAV에서 정치적 이데올로기 성향에 따라 외국인 이주자 관련정책에 대한 종합적 의견(총점)에 차이가 나타나는지를 검정하시오. 그 발견을 전달하기 위해 표를 만들고 시각화를 함께 수행하시오. 그 결론을 간단히 정리하시오.



연습 3. PROTEST.SAV에서 정치참여도를 나타내는 합성지수를 구성하고 이것이 연도별로 어떻게 변화하였는지 살펴보기 위해 유의성 검정을 수행하시오. 그 발견을 전달하기 위해 표를 만들고 시각화를 함께 수행하시오.

