

# 교차분석

- 두 개의 질적변수 사이에 관련성이 있는지 검정하는 분석

ex) 지역, 사교육 여부의 관계

동일성 검정	독립성 검정
한 변수(지역)의 각 그룹(a,b 지역)에서 다른 변수(사교육 여부)의 그룹(시킴, 안시킴)에서 나타나는 비율이 동일한지 검정 ex) a지역의 시킴 비율과 b지역의 시킴 비율이 동일한지	두 변수, 두 질적변수가 서로 영향을 미치는지 검정

귀무가설 - 각 지역마다의 사교육 여부는 동일하다  
대립가설 - 각 지역마다 사교육 여부는 동일하지 않다.

[지역에 관계없이 100명을 뽑아 조사한 결과]

(표1) 관측빈도 표

지역	사교육여부		전체
	시킴	아니다	
a	18(64.3)	10(35.7)	28
b	20(58.8)	14(41.2)	34
c	20(52.6)	18(47.4)	38
전체	58(58)	42(42)	100

- 현재 각 지역마다의 비율에 차이를 보이고 있음
  - 그러나 해당 수치만을 비교해서 "각 지역마다 사교육 여부는 다르다"라고 말하면 안됨!!
    - 현재는 임의의 100을 조사한 결과임 - 다른 100명을 조사하면 같게 나올 수도 있으니까

## 귀무가설이 참이라고 가정하고 생각

- 모든 지역이 58%로 같다고 하고 싶다.
  - 만약 그렇다면, 각 지역의 예상 인원은 어떻게 될 것인가?

(표2) 기대빈도 표

지역	사교육여부		전체
	시키고 있다	아니다	
a	16.2	11.8	28
b	19.7	14.3	34
c	22	16	38
전체	58	42(42)	100.0

각 지역의 인원수에 기대되는 시킴여부를 곱한다

ex)  $28(a\text{지역 인원수}) * 58(\text{시키고 있는 인원수}) / 100(\text{전체 인원수}) = 16.2$  (a지역에서 사교육을 시키고 있다고 기대되는 인원수) = 기대빈도

## 카이제곱 분석

실제로 관찰된 빈도와 (귀무가설이 참일 때) 기대되는 빈도를 나타내는 두 표의 차이를 구한 값 모든 경우의  $(\text{관측빈도} - \text{기대빈도})^2 / \text{기대빈도}$  를 합해서 구한다.

- 큼 ; 관측빈도, 기대빈도 간에 차이가 많이 남 - 귀무가설 기각
- 작음 : 이 둘 간에 차이 적음 - 귀무가설 채택

즉, 이것의 차이에 따라 귀무, 대립가설 중 어떤 것을 택할지 결정하게 해줌

그럼 어느정도를 크다고 할 수 있는가?

## spss 돌려방

지역은 1,2,3, 여부는 1,2로 표기 돼있다고 하자.

ex)

.	지역	사교육여부
	1	1

이렇게 된다면 해당 자료는 a지역의 시키고 있다에 해당되는 것

1. 교차분석 tab - 지역은 행, 사교육 여부는 열로 이동시킴



2. 그 중, 셀 tab - 빈도(관측, 기대), 퍼센트(행, 열, 전체)의 출현여부 설정



3. 이러면 교차표만 작성이 되고 카이제곱 값은 계산되지 않음



- 지역 중 % = a지역 전체 사람 중, 시키고 있다가의 비율로 64.3%
- 사교육여부 중 % = 시키고 있는 전체 58명 중, a지역에 해당되는 사람의 비율 31.0%
- 전체 중 % = 전체 100중 18%

본인의 용도에 따라 어떤 퍼센트를 확인할 것인지 생각하면 됨.

그러나 보통 행 퍼센트(지역 중 %)를 사용하는 것이 유용하다.

## 카이제곱 값 얻기

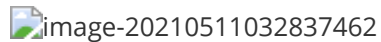
1. 통계량 tab



2. 행 퍼센트만 선택하여 실행하면 다음과 같다



3. 카이제곱 값은



- 카이제곱 값 = 0.913, 유의확률 값 = 0.633
- 유의 수준 0.05(5%)에서 검정하면 귀무가설을 채택하게 됨 - 카이제곱 값이 크게 차이가 난다고 할 수 없다.는 결론

**통계량의 자유도** = (행의 수 - 1)\*(열의 수 - 1)

- 전체 6개의 셀 중 2개 셀의 빈도에 의해 결정이 된다고 한다.