

카이-제곱 검정

영향을 주는 변수	영향을 받는 변수	분석 방법
범주형 척도의 자료	범주형 척도의 자료	카이제곱 검정
범주형 척도의 자료	연속형 척도의 자료	T-검정, ANOVA 분석
연속형 척도의 자료	범주형 척도의 자료	로지스틱 회귀 분석
연속형 척도의 자료	연속형 척도의 자료	회귀분석 구조방정식

범주형 척도의 자료가 범주형 척도의 자료에 영향을 받을 때 사용한다.

카이제곱검정 - 설문지 구성

- ✓ 출신 고등학교 > <
①서울고등학교 ②경기고등학교
2. 진학 대학교
① 서울대학교 ②연세대학교 ③고려대학교

둘 다 범주형 자료이고 정확한 수치로 비교할 수 없다. 물론 우리가 비교를 하기는 하지만 통계적으로는 그렇다는 뜻이다.

카이제곱검정 - 설문지 예시

예시 1.

1. 서울고 진학 대학교
서울대 10명 연세대 5명 고려대 5명

2. 경기고 진학 대학교
서울대 8명 연세대 6명 고려대 6명

예시 2.

1. 서울고 진학 대학교
서울대 10명 연세대 5명 고려대 5명

2. 경기고 진학 대학교
서울대 2명 연세대 2명 고려대 16명

	서울대	연세대	고려대
서울고 20	50% 10	25% 5	25% 5
경기고 20	40% 8	30% 6	30% 6

→ 두 고등학교 간 진학 대학교 비율이 거의 비슷하다.

→ 통계적인 차이가 있다고 하기 힘들다

	서울대	연세대	고려대
서울고	50%	25%	25%
경기고	10% +	10%	80%

→ 두 고등학교 간 진학 대학교 비율이 크게 차이가 있다.

→ 통계적인 차이가 있다고 할 수 있다

카이-제곱 검정에는 값과 유의확률이 있다. SPSS로 돌리면 아래와 같이 나온다고 한다.

사례 1 카이제곱 검정				사례 2 카이제곱 검정			
	값	자유도	정근 유의 확률 (양측 검정)		값	자유도	정근 유의 확률 (양측 검정)
Pearson 카이제곱	.404*	2	.817	Pearson 카이제곱	12.381*	2	.002
우도비	.405	2	.817	우도비	13.210	2	.001
선형 대 선형결합	.316	1	.574	선형 대 선형결합	11.363	1	.001
유효 케이스 수	40			유효 케이스 수	40		

모든 산포도에는 대표값이라는게 있다. 평균, 중앙값이 그것이다. 그 대표값에서 먼 값이 많으면 그 분포가 산포가 크다고 한다. 카이제곱이라는건 분산과 비슷한 개념이다.

카이제곱이 크면 두 범주형 척도(고등학교-대학교) 간의 차이가 많다는 뜻이 된다. 12.381 쪽에서 차이가 크다고 말할 수 있다. 보통 카이제곱의 값보다는 유의확률을 주로 본다.

유의확률이 작으면 차이가 있다는 뜻이다.

유의확률이 크면 차이가 없다는 뜻이다.

유의확률의 기준은 0.05이고 이것을 유의수준이라고 한다. 0.05보다 작으면 차이가 있다고 본다.