

# 가설검증을 이용한 분석 실습 - 교차분석(카이제곱 검정)

•[분석실습] 교차 분석을 알아봅니다.

공부함	공부안함	합격	불합격
1	0	1	0
1	0	1	0
0	1	0	1
0	1	0	1
1	0	1	0
1	0	1	0
0	1	1	0
0	1	1	0
0	1	0	1
1	0	0	1
1	0	0	1
1	0	1	0
1	0	1	0

	시험		계
	합격	불합격	
공부함	18	7	25
공부안함	12	13	25
계	30	20	50

빈도표

위 네모 부분이 교차빈도  
(관찰빈도)임

**H1(대립가설) : 공부를 열심히 하면 시험에 합격한다.**

**H0 (귀무가설): 공부와 시험은 관련이 없다.**

## 가설검증을 이용한 분석 실습 - 교차분석

카이제곱 검정은 관찰빈도와 기대빈도의 차이를 구하는 건데 만약 그 결과값이 차이가 없다면 두 변수는 서로 관계없이 따로 따로 존재하는 거고 차이가 클수록 연관성이 크다고 생각하면 됩니다.

	시험		계
	합격	불합격	
공부함	15	10	25
공부안함	15	10	25
계	30	20	50

기대 빈도는 각 교차빈도의 행과 열에 있는 주변 빈도값을 곱해서 구함.

# 가설검증을 이용한 분석 실습 - 교차분석

## •카이제곱 검정은

범주형 자료가 범주형 자료에 어떤 영향을 미치는지 분석할 때 사용되는 카이제곱검정에 대해서 알아보도록 하겠습니다.  
영향을 주는 변수, 영향을 받는 변수 모두가 범주형 자료라는 것입니다.

영향을 주는 변수	영향을 받는 변수	통계분석방법
범주형 자료	범주형 자료	카이제곱 검정
	연속형 자료	T검정 분산분석
연속형 자료	연속형 자료	회귀분석 구조방정식
	범주형 자료	로지스틱 회귀분석

# 가설검증을 이용한 분석 실습 - 교차분석

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

○ O 값은 관찰빈도 값, E 값은 기대빈도 값

	시험		
	합격	불합격	계
공부함	18	7	25
공부안함	12	13	25
계	30	20	50

	시험		
	합격	불합격	계
공부함	15	10	25
공부안함	15	10	25
계	30	20	50

$$\chi^2 = \frac{(18-15)^2}{15} + \frac{(7-10)^2}{10} + \frac{(12-15)^2}{15} + \frac{(13-10)^2}{10}$$

$$= 0.6 + 0.9 + 0.6 + 0.9$$

$$= 3 \leftarrow \text{우리가 구한 카이제곱 값입니다.}$$

# 가설검증을 이용한 분석 실습 - 교차분석

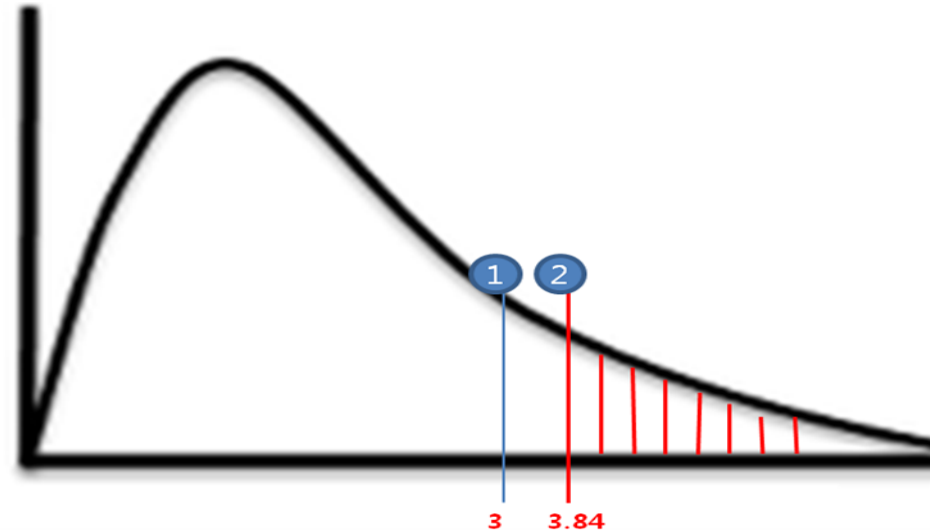
카이 제곱 값을 카이 제곱 표를 보고 구해야 하는데 자유도값이 필요함.  
 자유도는  $(x \text{ 범주 수} - 1) \times (y \text{ 범주 수} - 1)$  의 식으로 구함.  
 이번 예제에서는 x 범주 값이 공부함/공부안함 으로 2 가지 였고 y 범주값이 합격/불합격 2 가지 였으므로 자유도 값은 1 이 나옴.

$\alpha$ v	0.995	0.99	0.975	0.95	0.9	0.5	0.1	0.05	0.025	0.01	0.005
1	0.00004	0.0002	0.001	0.004	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1										
9	1										
10	2										
11	2										
12	3										
13	3										
14	4										
15	4										
16	5										
17	5										
18	6										
19	6										
20	7										
21	8										
22	8										
23	9										
24	9										
25	10										
26	11										
27	11										
28	12	12.46	13.56	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28
29	13	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59
30	13	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89
40	20	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69
50	27	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15
60	35	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38
70	43	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.43
80	51	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33
90	59	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.15	118.14	124.12
100	67	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81

유의 수준을 5% 정도로 하고 신뢰수준을 95% 정도로 하니까 자유도 1 과 유의수준 0.05 가 교차하는 지점의 값은 3.84 로 측정됨.

카이 제곱표에서 값이 3.84가 나왔다는 것의 의미는 카이 제곱 그래프에서 3.84 보다 적은 값의 카이 제곱 값은 귀무 가설이 참이 되고 크면 대립가설이 참이 된다는 뜻임.

## 가설검증을 이용한 분석 실습 - 교차분석



위 그림에서 2번 오른쪽 구간이 유의수준 5%에 속하는 구간임.  
카이제곱을 구한 값이 3.84보다 크면 대립 가설이 성립이 되고 3.84보다 작은 값이 구해지면 귀무가설이 성립 되므로 계산으로 구한 카이제곱 값은 3이므로 대립 가설 구간에 속하는 것이 아님. 즉 귀무 가설이 성립된다는 뜻이고 그 말은 공부하는 것과 시험에 합격하는 것은 상관이 없다는 결론을 낼 수 있다는 의미임.

# 가설검증을 이용한 분석 실습 - 교차분석

```
> install.packages("gmodels")
```

Installing package into 'C:/Users/seojinsu/Documents/R/win-library/3.1'

(as 'lib' is unspecified)

--- 현재 세션에서 사용할 CRAN 미러를 선택해 주세요 ---

URL 'http://cran.nexr.com/bin/windows/contrib/3.1/gmodels\_2.15.4.1.zip'을 시도합니다

( 지면 관계상 설치 과정은 생략합니다)

다운로드된 바이너리 패키지들은 다음의 위치에 있습니다

C:\Users\seojinsu\AppData\Local\Temp\RtmpMLrqEA\downloaded\_packages

```
> library(gmodels)
```

## 가설검증을 이용한 분석 실습 - 교차분석

```
> study <- read.csv("pass_cross.csv",header=T)
```

```
> study
```

	공부함	공부안함	합격	불합격
1	1	0	1	0
2	1	0	1	0
3	0	1	0	1
4	0	1	0	1
5	1	0	1	0
6	1	0	1	0

( 지면 관계상 이하 내용은 생략합니다. 총 50 건 입니다 )



## 가설검증을 이용한 분석 실습 - 교차분석

```
> CrossTable(study$공부합, study$합격, chisq=T)
```

Cell Contents

	N	Chi-square contribution	N / Row Total	N / Col Total	N / Table Total
study\$공부합					
0	13	0.900	0.520	0.260	0.260
1	7	0.280	0.350	0.140	0.140
Column Total	20	0.400	0.400	0.600	0.600

Total Observations in Table: 50

study\$공부합	study\$합격	0	1	Row Total
0	0	13	12	25
0	1	0.900	0.600	0.500
1	0	0.280	0.720	0.500
1	1	0.350	0.600	0.600
1	2	0.140	0.360	0.500
Column Total	0	20	30	50
Column Total	1	0.400	0.600	1.000

아래 표의 각 행에 대한 설명입니다.

첫번째 줄은 해당 조건의 데이터 건수이고

두번째줄은 해당 조건의 카이제곱 값이고

세번째 줄은 전체 조건대비 해당 행의 비율이고

네번째 줄은 전체 조건대비 해당 컬럼의 비율이고

마지막줄은 전체 데이터 대비 해당 값의 비율임.

## 가설검증을 이용한 분석 실습 - 교차분석

study\$공부함	study\$합격		Row Total
	0	1	
0	1 → 13	12	25
	2 → 0.900	0.600	
	3 → 0.520	0.480	0.500
	4 → 0.650	0.400	
	5 → 0.260	0.240	
1	7	18	25
	0.900	0.600	
	0.280	0.720	0.500
	0.350	0.600	
Column Total	20	30	50
	0.400	0.600	

- 1 번은 공부안하고(0) , 합격못한(0) 사람 수 13 건 입니다.
- 2 번은 앞에서 관찰빈도와 기대빈도값을 사용해서 구한 13에 대한 카이제곱 값이고
- 3 번은 해당 Row Total 값이 25 에 대한 13 의 빈도수 52% 이고
- 4 번은 해당 컬럼의 Total 값 20 에 대한 13 의 빈도수 65% 이고
- 5 번은 테이블 전체 50 에 대한 13의 빈도수 26% 입니다.

## 가설검증을 이용한 분석 실습 - 교차분석

```
Statistics for All Table Factors
```

```
Pearson's Chi-squared test
```

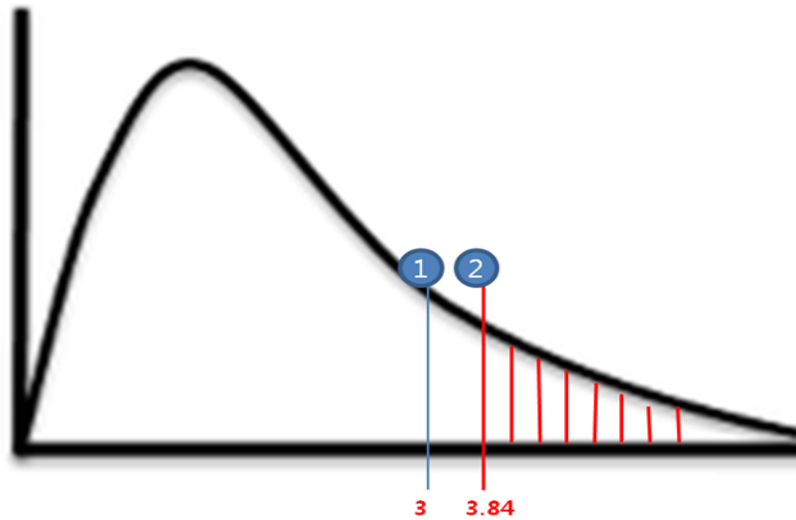
```
-----  
Chi^2 = 3      d.f. = 1      p = 0.08326452
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
-----  
Chi^2 = 2.083333      d.f. = 1      p = 0.1489147
```

CrossTable( ) 명령을 줄 때 마지막에 chisq=T 라는 옵션을 주면 보다 세부적인 결과까지 출력함.

# 가설검증을 이용한 분석 실습 - 교차분석



Statistics for All Table Factors			
Pearson's Chi-squared test			
Chi^2 = 3	d.f. = 1	p = 0.08326452	
Pearson's Chi-squared test with Yates' continuity correction			
Chi^2 = 2.083333	d.f. = 1	p = 0.1489147	