

#### 4. 독립성(Independence)검정과 동질성(Homogeneity)검정

- 독립성 : 두 변수가 서로 연관성을 가지고 있는지에 대한 검정  
학생들의 영양상태와 성적간에 관계  
취득학위와 소득수준간의 관계
- 동질성 : 변수들의 분포가 동일한지 살펴보는 검정  
폐암 발병율이 흡연자와 비흡연자간에 차이가 있는가?  
A/B 지역의 AIDS 감염비율이 차이가 있는가?

⇒ 서로 다른 개념이지만 동일한  $\chi^2$ -검정(chi-square test)를 수행

##### 1) 독립성/동질성 $\chi^2$ -검정의 개념

변수 1	변수 2				total
	1	2	...	c	
1	$O_{11}$	$O_{12}$	...	$O_{1c}$	$n_{1.}$
2	$O_{21}$	$O_{22}$	...	$O_{2c}$	$n_{2.}$
:	:	:	:	:	:
r	$O_{r1}$	$O_{r2}$	...	$O_{rc}$	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.c}$	$n_{..}$

$P(AB) = P(A)P(B)$ 의 원리를 이용함.

변수 1과 변수 2가 독립이라면,

$$O_{11} \text{의 기대값 } E_{11} = \left(\frac{n_{1.}}{n_{..}}\right) \cdot \left(\frac{n_{.1}}{n_{..}}\right) \cdot n_{..}$$

⇒ 모든 셀에 대해 계산이 가능함.

## 2) 검정통계량

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{\alpha}((r-1)(c-1))$$

## 3) 기각역

$$\chi^2 > \chi^2_{\alpha}((r-1)(c-1))$$

## 4) 예제

성별	전공			Total
	A 전공	B전공	C전공	
남자	75	46	23	144
여자	30	32	24	86
Total	105	78	47	230

학부 1학년생 230명을 대상으로 남녀별 전공 선택(3개전공)에 차이가 있는지 알아보려고 조사하여 위와 같은 교차표를 얻었다.

## - 가설

귀무가설 : 성별과 전공선택과는 관계가 없다. (서로 독립이다)

대립가설 : 성별에 따라 전공선택에 차이가 있다.

(서로 독립이 아니다 )

- 검정통계량

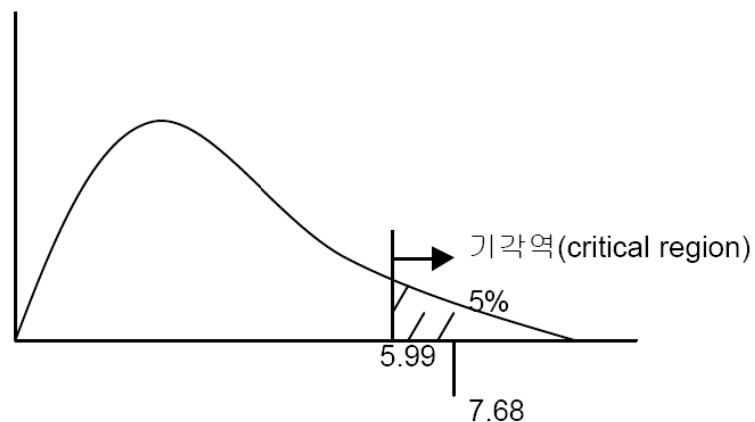
$$E_{11} = \left( \frac{105}{230} \right) \left( \frac{144}{230} \right) 230 = 65.7$$

$\vdots$

$$E_{23} = \left( \frac{86}{230} \right) \left( \frac{47}{230} \right) 230 = 17.6$$

$$\chi^2 = \frac{(75 - 65.7)^2}{65.7} + \dots + \frac{(24 - 17.6)^2}{17.6} = 7.68$$

$$\sim \chi^2_{\alpha}(df = (2 - 1)(3 - 1) = 2)$$



- 결론

성별에 따라 전공선택은 다르다.

- 가설은 양측검정인데, 왜 기각역은 한쪽방향만 있을까?

⇒ 검정통계량이 작다는 것은 관찰치와 기대값의 차이가 없음.

⇒ 두 변수간의 독립임을 의미함.

⇒ 표준정규분포를 따르는 확률변수  $Z$ 의 제곱은 자유도 1인  $\chi^2$  분포를 따른다.