

Chapter04

통계분석 II

I 카이제곱 검정

M T W T F S S

FASTCAMPUS
ONLINE

금융공학/퀀트 I

강사. 장순용

I 키포인트

- 카이제곱 검정.
- 도수분포표와 적합도 검정.
- 분할표와 독립성 검정.

I 도수분포표 (Frequency table)

- 도수분포표는 한개의 **명목형(범주형)** 변수가 있을 때 유형의 도수 (거듭된 횟수)를 집계한 표이다.
- 상대도수분포표는 도수의 상대적인 비율을 나타내어 준다.
- 연속형 변수를 사용하는 경우에는 구간을 설정해서 명목형 변수와 비슷하게 만들어 준다.
- 위의 경우 변수의 구간을 계급이라고 부른다.

I 분할표 (Contingency table)

- 분할표는 **두 개의 명목형(범주형)** 변수의 도수를 정리한 표이다.
- 분할표에서 도수분포표를 쉽게 구할 수 있다 (margin).

I 단변량 카이제곱검정 (Pearson's Chi squared test)

- 도수분포와 모형과의 차이 여부를 밝히기 위해서는 단변량 카이제곱검정을 실시한다:
 - 귀무가설 H_0 : 도수분포가 제시된 모형과 다르지 않다.
 - 대립가설 H_1 : 도수분포가 제시된 모형과 다르다.
- 적합도 검정 (goodness of fit test)라고도 부른다.
- 두개의 유형만이 있는 경우 \Rightarrow 비율검정 (proportion test) 적용.

I 단변량 카이제곱검정 (Pearson's Chi squared test)

- 다음과 같은 검정 통계량을 사용한다:

$$\text{검정 통계량} = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

- $E_i = np_i$ 는 모형이 제시하는 수치이고 실제 측청된 값은 O_i 이다.
- 위 검정 통계량은 자유도가 $k - 1$ 인 카이제곱 분포를 따른다.
 - O_i 의 합은 n 과 같다는 제약 조건이 있음.
 - 제약 조건은 자유도를 감소시킴: $k-1$.

I 분할표 카이제곱검정 (Chi squared test in two way table)

- 분할표를 사용해서 두 변수 사이의 독립성을 검정하고자 한다.
 - 귀무가설 H_0 : 두 명목형 변수는 독립적이다.
 - 대립가설 H_1 : 두 명목형 변수는 독립적이지 않다.
- 독립성 검정 (independence test)라고도 부른다.

I 분할표 카이제곱검정 (Chi squared test in two way table)

- 다음과 같은 검정 통계량을 사용한다: $r =$ 행의 수, $c =$ 열의 수 이다.

$$\text{검정 통계량} = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- 위 검정 통계량은 자유도가 $(r - 1) \times (c - 1)$ 인 카이제곱 분포를 따른다.

→ 행 방향의 합과 열 방향의 합은 고정적임. 이것은 제약 조건

I 분할표 카이제곱검정의 필요성

	남 (M)	여 (F)	계
A반	30	70	100
B반	50	40	90
계	80	110	190

- 반의 확률만 독립적으로 생각해 보면 다음과 같다:

$$P(\text{반} = A) = \frac{100}{190}, \quad P(\text{반} = B) = \frac{90}{190}$$

- 성별의 확률만 독립적으로 생각해 보면 다음과 같다:

$$P(\text{성별} = \text{남}) = \frac{80}{190}, \quad P(\text{성별} = \text{여}) = \frac{110}{190}$$

I 분할표 카이제곱검정의 필요성

	남 (M)	여 (F)	계
A반	30	70	100
B반	50	40	90
계	80	110	190

- 만약에 성별과 반이 서로 독립적인 상황이라면 다음과 같은 인수분해가 가능해 진다:

$$\begin{aligned}
 P(\text{반}, \text{성별}) &= P(\text{반}) \times P(\text{성별}) \\
 \Rightarrow N(\text{반}, \text{성별}) &= P(\text{반}) \times P(\text{성별}) \times N_{tot} \\
 &= E_{ij} \quad \text{“변수가 독립인 경우의 기대값”}
 \end{aligned}$$

FAST CAMPUS ONLINE • 그러므로 기대값 E_{ij} 를 실제 측청된 값 O_{ij} 와 비교하여 검정할 필요가 있다.

장순용 강사.

I 끝.

감사합니다.

