카이제곱 독립성 검정(chi-square independence test)

❑ 분석기법에 대한 정의

- 두 개 이상의 명목형 변수 간의 동질성(homogeneity), 연관성(association) 및 독립성(independence)을 검증하는 분석기법
- 예를 들어, 성별에 따른 재발여부에 대한 차이검증 또는 치료경험여부에 따른 병기에 대한 독립성 검증을 수행할 때 카이제곱 독립성 검정을 진행
- 특정 명목형 변수에 따라 다른 명목형 변수의 응답비율이 동일하다는 가설을 검증하는 분석기법

❑ 카이제곱 통계량에 대한 정의

- Pearson's Chi-square =

$$\sum \frac{(\text{관측도수} - \text{기대도수})^2}{(\text{기대도수})}$$

- 카이제곱 검정통계량은 (관측도수 – 기대도수)의 제곱값을 기대도수로 나눠준 값이며, 관측도수와 기대도수의 차이가 클수록 카이제곱 통계량은 커진다.
- 기대도수는 두 명목형 변수가 독립일 때의 기대되는 도수를 의미함.

❑ 카이제곱 통계량 사용시의 주의점

- 일반적으로 자유도가 1 인 2×2 분할표 검정에서는 Yates 의 연속성 수정을 하는 것이 바람직함.

$$\sum \frac{\left(|O - E| - \frac{1}{2}\right)^2}{E}$$

(O = 관측도수, E = 기대도수)

- 카이제곱 검정에서 각 칸의 기대도수가 5 이상이어야 하며, 만약 5 이하인 경우 Fisher 의 정확도 검정(Fisher's exact test)을 사용하는 것이 좋음
- 각 관측치 간에는 독립성이 만족되어야 하며, 독립성이 만족하지 않을 때는 McNemar 검정을 생각할 수 있음

#1422, Officetel 3 Danji, Gyeonghuigungeui Achim
72 Naesu-dong, Jongno-gu, Seoul 110-872, South Korea
Tel_+82-2-6271-7140~1/ **Cellular Phone**_+82-10-2171-7140
URL_ www.stateer.com

❑ 기대확률 산정방법

- 남자일 기대확률은 1/2, 합격일 기대확률은 1/2 이며, 남자이면서 합격일 기대확률은 (1/2)*(1/2) = 1/4 로 정의됨.
- 기대확률은 두 명목형 변수가 독립임을 가정했을 때의 특정 셀에 포함될 확률을 의미함.

❑ 연관성 측도

- **파이계수(Pearson's phi coefficient) = φ = sqrt(chi-square/n)**
  - → n 은 전체 표본의 크기
  - → 카이제곱 통계량을 기준으로 생성된 통계량으로 카이제곱 통계량보다 작은 값을 가지게 됨.
- **분할계수(contingency coefficient) = P = sqrt[φ^2/ (φ^2+1)]**
  - → 파이계수를 이용하여 계수값이 0 에서 1 사이의 값을 갖도록 조절한 통계량
  - → 1 에 가까울수록 유의하며, 0 에 가까울수록 두 변수가 독립임을 의미함
- **크레머의 V(Cramer's V) = V = sqrt[φ^2/ min{(I – 1), (J – 1)}]**
  - → 분할계수는 정확히 1 의 값을 가지는 경우가 없으므로 파이계수의 상한값(upper bound)으로 파이계수를 나누어 계수의 범위가 0 에서 1 사이의 값을 가지도록 조정한 통계량
  - → I 와 J 는 행과 열의 크기를 의미함

#1422, Officetel 3 Danji, Gyeonghuigungeui Achim
72 Naesu-dong, Jongno-gu, Seoul 110-872, South Korea
Tel_+82-2-6271-7140~1/ **Cellular Phone**_+82-10-2171-7140
URL_ www.stateer.com

❑ 카이제곱 통계량 이해를 위한 예시(1/2)

➢ 카이제곱 통계량 = $(40-50)^2/50 + (60-50)^2/50 + (60-50)^2/50 + (40-50)^2/50 = 8$

변수	구분	합격	불합격	전체
성별	남	40(관측도수) 1/4(기대확률) 50(기대도수)	60 1/4 50	100 1/2
	여	60 1/4 50	40 1/4 50	100 1/2
전체		100 1/2	100 1/2	200

❑ 카이제곱 통계량 이해를 위한 예시(2/2)

➢ 카이제곱 통계량 = $(20-50)^2/50 + (80-50)^2/50 + (80-50)^2/50 + (20-50)^2/50 = 72$

변수	구분	합격	불합격	전체
성별	남	20(관측도수) 1/4(기대확률) 50(기대도수)	80 1/4 50	100 1/2
	여	80 1/4 50	20 1/4 50	100 1/2
전체		100 1/2	100 1/2	200

#1422, Officetel 3 Danji, Gyeonghuigungeui Achim
72 Naesu-dong, Jongno-gu, Seoul 110-872, South Korea
Tel_+82-2-6271-7140~1/ **Cellular Phone**_+82-10-2171-7140
URL_ www.stateer.com