는 **상관분석**과 **교차분석**(crosstabs)이 있다. 분석하려는 변수가 구간척도 또는 비율척도로 측정된 변수인 경우에는 상관분석으로 연관성 정도를 분석하는 반면 명목척도나 서열척도로 측정된 범주변수인 경우에는 **분할표**를 이용한 교차분석으로 변수 간의 독립성 여부를 분석하게 된다.

수많은 양적 자료는 두 개 이상의 **범주변수**(categorical variable) 또는 분류기준에 따라 분류할 수 있는데 이때 우리는 이러한 기준들이 서로 영향을 전혀 미치지 않는 독립적인가에 관심을 갖게 된다. 예를 들면 동일한 제품의 상표 선호가 성별과 독립적인가를 알고자 할 경우에는 표본으로 고객들을 추출하고 그들의 성별에 따른 상표 선호를 분류한다.

이때 두 개 이상의 독립변수가 관련되어 있기 때문에 사용되는 표가 **분할표**(contingency table)이다. 분할표는 모집단에서 추출된 표본자료를 한 분류기준에 따른 r개의 행과 다른 분류기준에 따른 c개의 열로 분류하여 작성한 통계표이다. 표의 각 칸(cell)은 행과 열의 교차점이다. 표의 각 칸에는 표본 가운데 해당하는 요소들의 도수를 기록한다.

검정을 위해서는 분할표의 **관찰도수**와 두 변수가 상호독립일 때의 **기대도수**를 비교하여야 한다. 만일 기대도수가 관찰도수와 상당한 차이를 보이면 검정통계량이 큰 값을 갖게 되어 두 변수가 독립적이라는 귀무가설을 기각하게 된다. 즉 기대도수와 관찰도수가 일치하면 두 변수는 독립적이라고 판단할 수 있다.

두 변수 간의 독립성을 전제로 할 때 각 칸의 기대도수를 구하는 공식은 다음과 같다.

$$E_{ij} = \frac{(\text{행 } i\text{의 합계})(\text{열 } j\text{의 합계})}{\text{표본크기}} = \frac{R_i\,C_j}{n}$$

R_i: i 번째 행의 합계
C_j: j 번째 열의 합계

귀무가설 H_0을 검정할 검정통계량은 다음과 같이 구한다.

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

검정의 임계치로 사용할 χ^2값을 구하기 위해서는 자유도를 결정해야 하는

데 $(r \times c)$분할표의 자유도는 다음과 같이 구한다.

$$df = (r-1)(c-1)$$

만일 계산된 χ^2값이 표로부터 구하는 유의수준 α와 자유도 $(r-1)(c-1)$의 χ^2값보다 크면 귀무가설을 기각하게 된다. 따라서 두 변수의 독립성에 관한 검정은 다음과 같이 정리할 수 있다.

두 변수의 독립성에 대한 우측검정

H_0: 두 변수는 독립적이다.

H_1: 두 변수는 종속적이다.

만일 $\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} > \chi^2_{(r-1)(c-1),\,\alpha}$이면 귀무가설 H_0을 기각

만일 p값$<\alpha$이면 귀무가설 H_0을 기각

예 13-2 다음은 무작위로 추출한 150명의 고객을 대상으로 성별에 따라 동일한 제품의 상표 선호에 차이가 있는지를 조사하기 위하여 수집한 자료이다. 성별과 상표 선호는 독립적인지 유의수준 5%로 검정하라.

성별 \ 상표	A	B	C	합 계
남	20	40	20	80
여	30	30	10	70
합 계	50	70	30	150

풀 이 H_0: 두 변수는 독립적이다.

H_1: 두 변수는 종속적이다.

$$E_{11} = \frac{80(50)}{150} = 26.667 \qquad E_{12} = \frac{80(70)}{150} = 37.333$$

$$E_{13} = \frac{80(30)}{150} = 16 \qquad E_{21} = \frac{70(50)}{150} = 23.333$$

$$E_{22} = \frac{70(70)}{150} = 32.667 \qquad E_{23} = \frac{70(30)}{150} = 14$$

$$\chi_C^2 = \frac{(20-20.667)^2}{26.667} + \frac{(40-37.333)^2}{37.333} + \frac{(20-16)^2}{16}$$