

1. 교차분석이란?

1) 의미

교차분석은 χ^2 분석 혹은 χ^2 을 의미하는 카이제곱 분석이라고도 불린다. 교차분석은 명목변수나 서열변수인 두 변수의 관련성 여부를 분석하는 것이다. 교차해서 YES가 나타날 확률을 분석하는 것이므로 교차분석이라는 이름이 붙었다.

예를 들어, K대학교 행정학과 학생들을 대상으로 성별과 종교의 관련성 여부를 분석한다고 하자. 성별은 남자 혹은 여자로 구분되는 명목변수이다. 그런데 종교도 명목변수이다. 이처럼 두 변수, 즉 성별과 종교의 관련성 여부를 통계적으로 분석하는 방법을 교차분석이라 한다.

2) 카이제곱 분포

교차분석은 카이제곱 분포를 활용한다. 카이제곱 분포는 k개의 서로 독립적인 변수를 각각 제공하여 그 제공한 값들을 합한 결과로 얻어지는 분포이다. 여기서 k는 서로 독립적인 정규분포를 따르는 확률변수이다. 자유도 k의 카이제곱 분포는 다음과 같다.

$$Q = \sum_{i=1}^k \chi_i^2$$

2. 검정통계량의 도출과 가설의 설정

1) 검정통계량의 도출

교차분석의 검정통계량은 χ^2 이다. χ^2 는 다음의 공식을 통해 계산된다.

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

여기서 O는 관측값, E는 기댓값이다.

2) 가설의 설정

교차분석의 귀무가설은 “명목변수인 x와 명목변수인 y는 관련성이 없다.”이고 대립가설은 “x와 y는 관련성이 있다.”이다. 검정통계량 χ^2 가 0에 가까워질수록 유의확률이 높아져서 귀무가설을 채택할 가능성이 커지므로 귀무가설은 $\chi^2=0$ 으로 정의한다. 따라서 대립가설은 $\chi^2 \neq 0$ 이다.

3. R 활용 교차분석

ad_layer와 multichild의 관련성 여부를 통계적으로 분석하고자 한다. 데이터는 이원분산분석에 활용된 데이터를 그대로 사용한다.

```
> setwd("C:/Users/USER/Desktop/교육원초급/data/xlsx 파일")
> library(openxlsx)
> mydata4 = read.xlsx("anova_two_way.xlsx",1) # 파일을 불러와 mydata4로 저
> attach(mydata4) # 불러온 데이터를 R과 연동
> View(mydata4) # 불러온 데이터를 R Studio에서 확인
> names(mydata4) # 불러온 데이터에 포함된 변수의 이름을 확인
[1] "cities"      "birth_rate" "ad_layer"   "multichild" "ID"
> table(ad_layer,multichild)
      multichild
ad_layer NO YES
자치구 64   5
자치군 78   4
자치시 69   6
> TAB = table(ad_layer,multichild)
> chisq.test(TAB)

Pearson's Chi-squared test
```

```
data: TAB  
X-squared = 0.67562, df = 2, p-value = 0.7133
```

table이라는 명령어를 통하여 교차분석 테이블을 출력할 수 있다. 카이제곱분석 결과 검정통계량이 0.67562이고, 유의확률이 0.7133으로 귀무가설이 채택되었음을 확인할 수 있다.