

1

다음은 랜덤하게 선택된 806명의 Facebook 사용자에게 대해 Facebook의 privacy setting 기능의 사용법을 알고 있는지, 에 대해 조사한 결과이다. 성별과 privacy setting의 사용법 인지 여부가 관계가 있다고 할 수 있는가? 유의수준 5%에서 이를 검정하시오.

성별 \ 사용법 인지여부	남	여	합계
알고 있다	288	378	666
잘 모른다	10	7	17
전혀 모른다	61	62	123
합계	359	447	806

* Survey USA, News Poll #17960, data collected February 16-17, 2011

```
x = np.array([[288, 10, 61], [378, 7, 62]])
table = pd.DataFrame(x, ["Male", "Female"], ["Know", "Not well", "Don't know"])
```

```
result = chi2_contingency(table)
chi, p, df, exp = result
print("chi2 = ", chi)
print("p value = ", p)
print("degree of freedom = ", df)
print("expected value = ")
print(exp)
```

'''

```
chi2 = 3.1290636322789274
p value = 0.07257003
degree of freedom = 2
expected value =
[[296.6426799 7.5719603 54.7853598]
 [369.3573201 9.4280397 68.2146402]]
```

'''

- 전체 수만이 주어졌기 때문에 이는 독립성 검정이다. 귀무가설은 서로 독립이다, 즉 각각의 조건(예: 사용법 인지)에 대한 다른 조건(예: 성별)의 비율이 똑같다는 것이다. 이에 따라 예상값을 각각 서로 독립인 경우의 비율로 둔다.
- 그 결과 p value는 0.21으로, 귀무가설이 채택된다. 예상값이 모두 5 이상이므로 그대로, 성별과 사용법 인지는 독립이라고, 유의수준 5%에서 유의미한 관계가 없다고 할 수 있다.

2

심폐 소생술은 심장과 폐의 활동이 갑자기 멈추었을 때 실시하는 응급처치법이다. 심폐 소생술 시행 시 혈액 멍침 등을 완화하기 위해 혈액 희석제(blood thinner)를 처방하는 경우가 있는데, 만약 심폐 소생술로 인해 신체 내부 손상을 입은 경우라면 혈액 희석제는 환자에게 치명적일 수도 있다. 다음은 심폐 소생술을 받은 환자 중 혈액 희석제를 사용하지 않은 환자들 50명과 사용한 환자들 40명을 대상으로 그들의 생존 여부를 조사한 결과이다.

	생존	사망	합계
혈액 희석제 사용안함	11	39	50
혈액 희석제 사용함	14	26	40
합계	25	65	90

혈액 희석제 사용 여부에 따른 생존률은 차이가 있는가? 유의수준 5%에서 이를 검정하시오.

```
x = np.array([[11, 39], [14, 26]])
```

```
table = pd.DataFrame(x, ["Blood thinner", "none"], ["Survive", "Dead"])
```

```
result = chi2_contingency(table)
chi, p, df, exp = result
print("chi2 = ", chi)
print("p value = ", p)
print("degree of freedom = ", df)
print("expected value = ")
print(exp)
```

```
'''
```

```
chi2 = 1.2800769230769231
p value = 0.25788473323931277
degree of freedom = 1
expected value =
[[13.88888889 36.11111111]
 [11.11111111 28.88888889]]
```

```
'''
```

- 합계가 주어진, 각 사용 여부에 따라 값을 모았기 때문에, 이는 동질성 검정에 해당한다. 따라서 귀무가설은 혈액 희석제를 사용하거나 사용하지 않았거나, 두 계열에서 생존과 사망의 비율이 같다는, 두 계열이 동질적이라는 것이다.
- 이를 카이제곱 검정을 통해 얻어보면, p value는 $0.26 > 0.05$ 이고, 예상값이 모두 5 이상이므로 그대로 귀무가설이 채택된다. 이에 따라 혈액희석제를 사용하거나 사용하지 않거나, 유의수준 5%에서는 생존과 사망 비율이 같다.

```
# add 1
```

자연대 학생 100명의 혈액형을 관찰하였다. 100명 중 A, B, O, AB형 학생수는 각각 다음과 같다.

혈액형	A	B	O	AB
인원수	17	33	35	25

우리나라 사람들의 혈액형별 비율은 A : B : O : AB = 2 : 3 : 3 : 2로 알려져 있다. 조사한 자연대 학생들의 혈액형별 비율은 우리나라 사람들의 혈액형별 비율과 일치하는지 알고 싶다.

- 1) 위 문제를 해결하기 위한 귀무가설과 대립가설을 설정하시오.
- 2) 유의수준 5%에서 위 실험 결과가 우리나라 사람들의 혈액형별 비율을 따라가는지 검정하시오. 이때 유의확률을 구하시오.

```
x = [17, 33, 35, 25]
```

```
y = [0.2, 0.3, 0.3, 0.2]
```

```
exp = [i*sum(x) for i in y]
```

```
result = chisquare(x, exp)
```

```
print(result)
```

```
'''
```

```
Power_divergenceResult(statistic=1.666666666666667, pvalue=0.6443698056370251)
```

```
'''
```

- 귀무가설은, 단순히 혈액형 비율이 A, B, O, AB 각각이 0.2, 0.3, 0.3, 0.2라는, 즉 각 인원수가 20, 30, 30, 20명이라는 것이다. 대립가설은 그렇지 않다는 뜻이 된다.
- 이를 chisquare 함수로 분석하면, p value는 0.64>0.05로, 유의수준 5%에서 실험 결과가 우리나라 사람들의 혈액형별 비율을 따라간다. 유의확률은 그대로 64%이다.

```
# add 2
```

임의로 20대와 30대 사람들 각각 100명을 뽑아 최근 1년 동안 응급실 가는 횟수를 조사하였다. 응급실을 간 횟수를 0번,1번,2번이상 3개의 카테고리로 분류해보니 20대가 응급실을 0번,1번,2번이상 간 경우는 각각 37,30,33명이었고 30대가 응급실을 0번,1번,2번이상 간 경우는 각각 35,25,40명이었다. 연령대별로 1년동안 응급실 가는 횟수가 차이가 있다고 할 수 있는지 유의수준 5%에서 확인하시오.

```
x = np.array([[37, 30, 33], [35, 25, 40]])
```

```
table = pd.DataFrame(x, ["20s", "30s"], ["0 time", "1 time", "2 times"])
```

```
result = chi2_contingency(table)
chi, p, df, exp = result
print("chi2 = ", chi)
print("p value = ", p)
print("degree of freedom = ", df)
print("expected value = ")
print(exp)
```

```
'''
```

```
chi2 = 1.1813338868133387
p value = 0.5539577030655446
degree of freedom = 2
expected value =
[[36. 27.5 36.5]
 [36. 27.5 36.5]]
```

```
'''
```

- 20대, 30대 각각 100명을 뽑았으므로 이는 동질성 검정이 된다. 이에 따라 귀무가설은 두 세대 간에서 응급실을 가는 빈도의 비율이 같다는 것이고, 대립가설은 그렇지 않다는 것이 된다.
- 가설 검정 결과, p value는 $0.55 > 0.05$ 로, 귀무가설이 채택된다. 예상값은 모두 5 이상이므로, 이에 따라 20대와 30대는 유의수준 5%에서, 서로 응급실을 가는 횟수의 분포에 차이가 있다고 할 수 없다.

```
# add 3
```

연령과 사용하는 핸드폰 브랜드와 연관이 있는지 알아보기 위해 200명을 대상으로 연령별로 사용하는 핸드폰의 브랜드를 조사하였다. 그 결과 20대는 삼성, 애플을 사용하는 사용자가 각각 38, 62명, 30대는 삼성, 애플을 사용하는 사용자가 각각 57, 43명이었다. 연령과 사용하는 핸드폰 브랜드는 서로 관계가 있는지 유의수준 5% 하에서 확인하시오.

```
x = np.array([[38, 62], [57, 43]])
```

```
table = pd.DataFrame(x, ["20s", "30s"], ["Samsung", "Apple"])
```

```
result = chi2_contingency(table)
```

```
chi, p, df, exp = result
```

```
print("chi2 = ", chi)
```

```
print("p value = ", p)
```

```
print("degree of freedom = ", df)
```

```
print("expected value = ")
```

```
print(exp)
```

```
'''
```

```
chi2 = 6.496240601503759
```

```
p value = 0.010810283451958084
```

```
degree of freedom = 1
```

```
expected value =
```

```
[[47.5 52.5]
```

```
 [47.5 52.5]]
```

```
'''
```

- 비록 20대와 30대 모두 100명이 나왔지만, 200명을 무작위로 뽑았으므로 이는 독립성 검정이 된다. 이에 따라 귀무가설은 연령대(20/30)와 사용하는 폰의 브랜드(삼성/애플)의 각 비율이 서로 독립이라는 뜻이고, 대립가설은 서로 종속이라는, 연관이 있다는 뜻이 된다.
- 가설 검정 결과, 각 예측값 모두 5 이상에서 p value가 $0.011 < 0.05$ 로, 유의수준 5%에서 귀무가설이 기각된다. 이에 따라 연령대와 사용하는 폰 브랜드는 서로 종속이고, 즉 서로 관계가 있다. 특히 20대가 30대보다 아이폰을 더 많이 사용한다는 점을 확인할 수 있다.