

Assignment #6
2016-19516, Sangjun Son

Example 1

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

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

```
1 facebook <- rbind(c(288, 278), c(10, 7), c(61, 62))
2 chisq.test (facebook)
```

```
-----
Pearson's Chi-squared test

data: facebook
X-squared = 0.5104, df = 2, p-value = 0.7748
-----
```

Explanation: 독립성 검정 모형, 남자가 Facebook의 privacy setting에 대해 알고 있을 확률, 잘 모를 확률, 전혀 모를 확률을 각각 p_{11}, p_{21}, p_{31} 라 하고, 여자에 대해서도 마찬가지로 p_{12}, p_{22}, p_{32} 라고 하면 검정하고자 하는 가설을 다음과 같다. 귀무가설 H_0 : 성별과 privacy setting의 사용법 인지 여부는 서로 독립이다와 대립가설 H_1 : $\neg H_0$. 독립성 검정 `chisq.test()`를 진행하였을 때 유의확률은 0.7748로 유의수준 5%에서 기각할 수 없다. 즉, 성별과 privacy setting의 사용법 인지 여부는 서로 독립이다.

Example 2

심폐 소생술은 심장과 폐의 활동이 갑자기 멈추었을 때 실시하는 응급처치방법이다. 심폐 소생술 시행 시 혈액 묽힘 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 환자들 중 혈액 희석제를 사용하지 않은 환자들 50명과 혈액 희석제를 사용한 환자들 40명을 대상으로 그들의 생존 여부를 조사 한 결과이다. 혈액 희석제 사용 여부에 따른 생존률은 차이가 있는가? 유의수준 5%에서 이를 검정하시오

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

```
1 blood.thinner <- rbind(c(11, 39), c(14, 26))
2 chisq.test (blood.thinner)
```

```
-----
Pearson's Chi-squared test with Yates' continuity
correction
-----
```

Assignment #6

2016-19516, Sangjun Son

```
data: blood.thinner
X-squared = 1.2801, df = 1, p-value = 0.2579
```

Explanation: 동질성 검정 모형, 혈액 희석제를 사용하지 않은 그룹과 사용한 그룹을 먼저 50명과 40명으로 지정 하였으므로 각 집단을 부차모집단으로 보고 혈액 희석제 사용 여부에 따른 생존률은 차이를 비교하면 된다. 이때 귀무가설 $H_0 : p_{1j} = p_{2j} (j = 1, 2)$ 과 대립가설 $H_1 : \neg H_0$ 로 설정 후 동질성 검정 `chisq.test()`를 진행한다. 자유도는 $df = (r - 1)(c - 1) = (2 - 1)(2 - 1)$ 임을 확인하고 유의확률은 0.2579로 유의수준 5%에서 기각할 수 없다. 즉, 혈액 희석제 사용 여부에 따른 생존률은 차이가 존재한다고 할 수 없다.

Example 3

3장 과제에서 사용하였던 iris 데이터셋을 분석하려고 한다. `data('iris')`의 코드를 실행하여 데이터셋을 불러올 수 있으며, R에 내장된 iris 데이터에는 붓꽃의 종별 꽃잎과 꽃받침의 크기가 저장되어 있다. 이 중 Sepal.Width 값이 3.3보다 크거나 같을 경우를 'L', 2.8보다 크거나 같거나 3.3보다 작은 경우를 'M', 2.8보다 작을 경우를 'S'로 분류할 때, setosa와 versicolor에 따른 붓꽃의 크기의 차이가 있다고 할 수 있는지 유의수준 5%에서 검정하여 보자.

```
1 data(iris)
2 setosa = iris[iris$Species == 'setosa',]
3 versicolor = iris[iris$Species == 'versicolor',]
4
5 setosa.L = setosa[setosa$Sepal.Width > 3.3,]
6 setosa.M = setosa[setosa$Sepal.Width >= 2.8 & setosa$Sepal.Width < 3.3,]
7 setosa.S = setosa[setosa$Sepal.Width < 2.8,]
8
9 versicolor.L = versicolor[versicolor$Sepal.Width > 3.3,]
10 versicolor.M = versicolor[versicolor$Sepal.Width >= 2.8 & versicolor$Sepal.Width < 3.3,]
11 versicolor.S = versicolor[versicolor$Sepal.Width < 2.8,]
12
13 iris.Sepal.Width = rbind(c(nrow(setosa.L), nrow(versicolor.L)),
14                          c(nrow(setosa.M), nrow(versicolor.M)),
15                          c(nrow(setosa.S), nrow(versicolor.S))); iris.Sepal.Width
16
17 chisq.test(iris.Sepal.Width)
```

Pearson's Chi-squared test

```
data: iris.Sepal.Width
X-squared = 49.116, df = 2, p-value = 2.161e-11
```

Explanation: 동질성 검정 모형, 예제 3과 동일한 방법으로 setosa의 집단과 versicolor 집단 내에서의 Sepal의 폭에 따라 크기를 나누었을 때의 그 비율이 동일한지에 대해 검정을 진행할 수 있다. 'L', 'M', 'S'에 대한 조건 대로 setosa와 versicolor 집단에 대해 slicing을 진행하고 각각의 개체의 개수를 저장하였다. 이렇게 구한 값을 바탕으로 분할표를 `rbind()`와 `c()`를 통해 `iris.Sepal.Width`에 저장하였다. 동질성 검정 `chisq.test()`를 진행하였고 유의확률은 2.16e-11로 유의수준 5%에서 기각할 수 있다. 즉, setosa와 versicolor에 따른 붓꽃의 크기의 차이가 존재한다.

Assignment #6
2016-19516, Sangjun Son

Example 4

vcd 패키지 안에 있는 Arthritis 데이터 안에는 각 환자별 정보와 투약 여부에 따른 상태 호 전 결과가 저장되어 있다. 데이터의 남성환자에 대해서, 투약 여부(Treatment)에 따른 결과(Improved)가 다르다고 할 수 있는지 유의수준 5%에서 검정하여 보자.

```
1 install.packages("vcd")
2 library("vcd")
3 data(Arthritis)
4 male = Arthritis[Arthritis$Sex == "Male",]
5 Arthritis.table = table(male$Treatment, male$Improved); Arthritis.table
6 chisq.test(Arthritis.table)
```

	None	Some	Marked
Placebo	10	0	1
Treated	7	2	5

Pearson's Chi-squared test

```
data: Arthritis.table
X-squared = 4.9067, df = 2, p-value = 0.086
```

Explanation: 동질성 검정 모형, 예제 3, 4과 동일한 방법으로 남성 중 투약 한 집단 Treated과 그렇지 않은 집단 Placebo 내에서의 Improved 결과에 대한 비율이 동일한지에 대해 검정을 진행할 수 있다. 먼저 남성 데이터를 추출하기 위해 Arthritis의 Sex 변수를 통해 slicing을 진행하고 table()을 이용해 분할표 생성을 하였다. chisq() 검정 결과 유의확률은 0.086으로 유의수준 5%에서 기각할 수 없다. 즉, 남성환자에 대해서, 투약 여부(Treatment)에 따른 결과(Improved)가 다르다고 할 수 없다.