



Ch10.교차분석(χ^2 분석)



연구방법의 이해

- 범주형자료(Categorical Data)와 분할표
 - 관측치들이 몇 개의 범주에 의해서 분류되어 각 범주에 해당되는 도수로 정리된 자료
 - 예) 성별, 연령별, 지역별 특성에 따라서 분석된 자료
 - 외국교육기관이 들어오는 것을 찬성, 반대
 - 연령에 따른, 시간대에 따른 시청자들의 시청 프로그램 선호도 비교
 - χ^2 분석법에 의해 분석



교차분석

- 범주형자료(Categorical Data)와 분할표
 - 관측치들이 몇 개의 범주에 의해서 분류되어 각 범주에 해당되는 도수로 정리된 자료
 - 예) 성별, 연령별, 지역별 특성에 따라서 분석된 자료
 - 외국교육기관이 들어오는 것을 찬성, 반대
 - 연령에 따른, 시간대에 따른 시청자들의 시청 프로그램 선호도 비교
 - χ^2 분석법에 의해 분석

실험설계 교차분석

■ 문제의 정의

- K 병원에서는 비타민과 감기와와의 관계를 연구하고자 한다.
- 감기가 걸리지 않은 사람을 대상으로 비타민을 투여할 실험군과 가짜약을 투여할 대조군으로 구분하고 겨울동안 감기가 걸렸는지를 확인하였다.
- 과연 비타민이 감기에 효과가 있었는지 검증해 보라
- (12.PreCRO.sav).

■ 가설

- 귀무가설(H_0) : 비타민처치와 감기발병은 관계가 없다.

$$H_0 : \pi_1 = \pi_2$$

- 연구가설(H_1) : 비타민처치와 감기발병은 관계가 있다.

$$H_1 : \pi_1 \neq \pi_2$$

실험설계 교차분석

■ 기대치 산출

$$E(x_1) = 50 \times \hat{\pi}_j = 50 \times \frac{45}{100} = 22.5$$

$$E(x_2) = 50 \times \hat{\pi}_j = 50 \times \frac{55}{100} = 27.5$$

■ 검정통계량

$$\chi_0^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} = \frac{(33 - 22.5)^2}{22.5} + \dots = 17.818$$

■ 검정결과

$\chi_0^2 = 17.818 > \chi^2(1) = 3.841$ 이므로 H_0 를 기각(p -value = 0.000)

실험설계 교차분석

	감기발병		계
	유	무	
비타민	17	33	50
Placebo	38	12	50

$$P_1 = \frac{17}{50} = 0.515$$

$$P_2 = \frac{38}{50} = 0.760$$

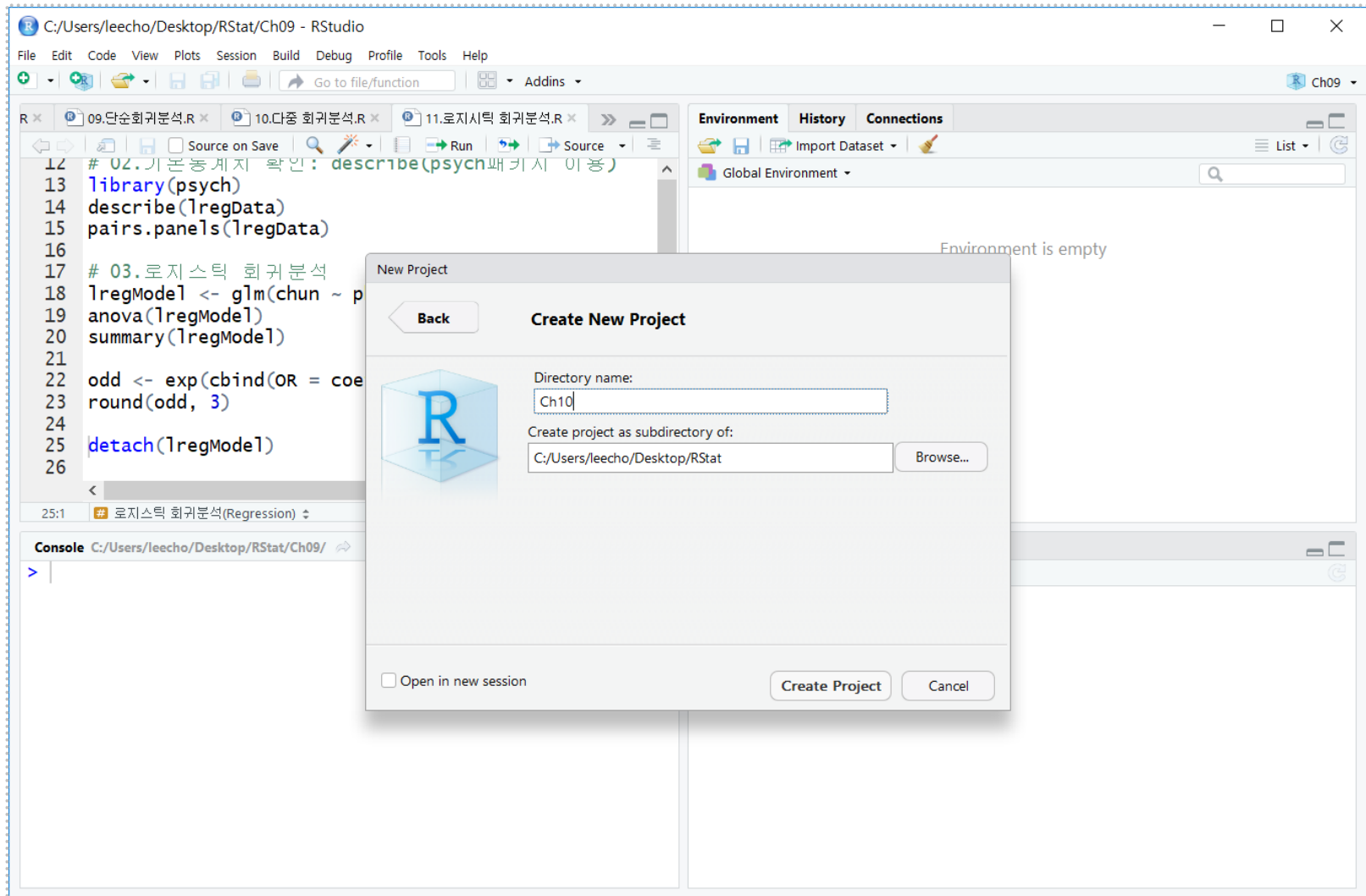
$$d = P_1 - P_2 = 0.515 - 0.760 = -0.255$$

* 소량의 아스피린을 복용하면 심근경색을 일으키는 사람이 25.5% 준다

$$RR = \frac{17/50}{38/50} = \frac{0.515}{0.760} = 0.678$$

* 상대위험율이 1보다 작으므로 아스피린을 복용했을 경우 안 복용했을때보다 심근경색을 일으킬 위험이 상대적으로 적다는 것을 의미

프로젝트 생성





실험설계 교차분석

- ##### 사전 교차분석: 동질성 (chi-square) #####
- # 01.데이터 불러오기
- precroData <- read.csv("./data/12.preCro.csv",
- header=TRUE,
- na.strings = ".")
- precroData\$group <- factor(precroData\$group,
- levels=c(1,2),
- labels=c("비타민","Placebo"))
- precroData\$cold <- factor(precroData\$cold,
- levels=c(1,2),
- labels=c("noCold","Cold"))

실험설계 교차분석

```
Console C:/Users/leecho/Desktop/RStat/Ch10/
> ##### 사전 교차분석: 동질성 (chi-square) #####
>
> # 01.데이터 불러오기
> precroData <- read.csv("./data/12.preCro.csv",
+                          header=TRUE,
+                          na.strings = ".")
> precroData$group <- factor(precroData$group,
+                             levels=c(1,2),
+                             labels=c("비타민", "Placebo"))
> precroData$cold <- factor(precroData$cold,
+                             levels=c(1,2),
+                             labels=c("noCold", "cold"))
>
> str(precroData)
'data.frame': 100 obs. of 2 variables:
 $ group: Factor w/ 2 levels "비타민", "Placebo": 1 1 1 1 1 1 1 1 1 1 ...
 $ cold : Factor w/ 2 levels "noCold", "cold": 1 1 1 1 1 1 1 1 1 2 ...
> |
```



실험설계 교차분석

- `attach(precroData)`
- #01.분할표 만들기
- `nTable <- table(group, cold)`
- `addmargins(nTable)`
- `pTable <- prop.table(nTable)`
- `addmargins(pTable)`

실험설계 교차분석

Console C:/Users/leecho/Desktop/RStat/Ch10/ ↗

```
> attach(precroData)
> #01. 분할표 만들기
> nTable <- table(group, cold)
> addmargins(nTable)
```

group	cold		Sum
	noCold	cold	
비타민	33	17	50
Placebo	12	38	50
Sum	45	55	100

```
> pTable <- prop.table(nTable)
> addmargins(pTable)
```

group	cold		Sum
	noCold	cold	
비타민	0.33	0.17	0.50
Placebo	0.12	0.38	0.50
Sum	0.45	0.55	1.00

```
> |
```

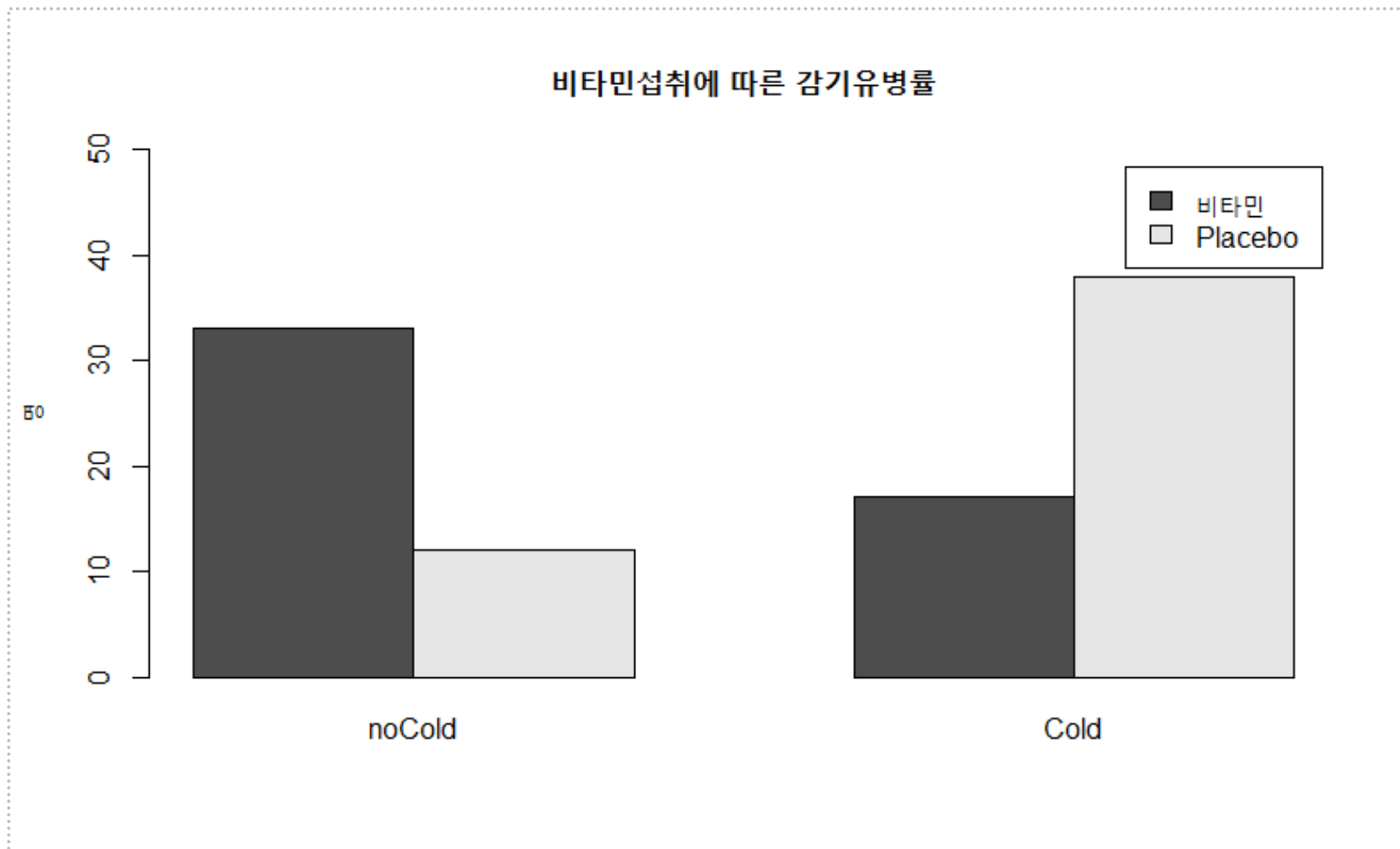


실험설계 교차분석

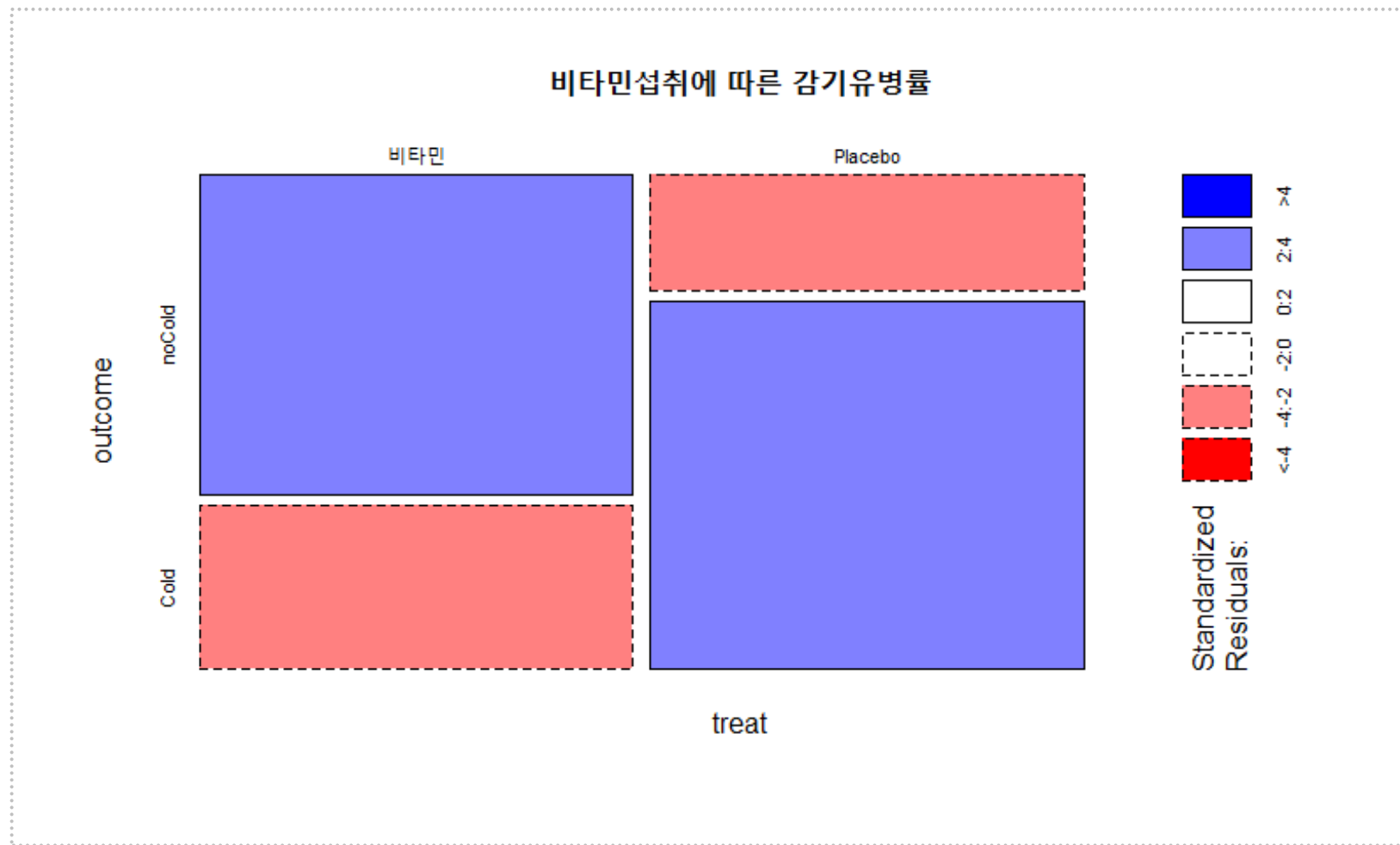
- #02.그래프 그리기
- `barplot(nTable, beside=TRUE,`
- `legend=TRUE, ylim=c(0, 50),`
- `ylab="명",`
- `main="비타민섭취에 따른 감기유병률")`

- # 모자이크 그래프
- `mosaicplot(nTable, shade=TRUE,`
- `xlab = "treat", ylab = "outcome",`
- `main="비타민섭취에 따른 감기유병률")`

실험설계 교차분석



실험설계 교차분석





실험설계 교차분석

- #03. 카이스케어 분석결과
- `chisq.test(nTable)`
- #04. 패키지 이용
- `install.packages("gmodels")`
- `library(gmodels)`
- `precroResult <- CrossTable(nTable,`
- `expected=TRUE,`
- `chisq=TRUE)`

실험설계 교차분석

Console C:/Users/leecho/Desktop/RStat/Ch10/ ↗

```
> #03. 카이스케어 분석결과  
> chisq.test(nTable)
```

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

data: nTable

X-squared = 16.162, df = 1, p-value = 5.816e-05

```
> library(gmodels)  
> precroResult <- CrossTable(nTable,  
+                             expected=TRUE,  
+                             chisq=TRUE)
```

Cell Contents

		N
	Expected N	
Chi-square	contribution	
	N / Row Total	
	N / Col Total	
	N / Table Total	

실험설계 교차분석

Console C:/Users/leecho/Desktop/RStat/Ch10/ ↗

	22.500	27.500	
	4.900	4.009	
	0.240	0.760	0.500
	0.267	0.691	
	0.120	0.380	
-----	-----	-----	-----
Column Total	45	55	100
	0.450	0.550	
-----	-----	-----	-----

Statistics for All Table Factors

Pearson's Chi-squared test

chi^2 = 17.81818 d.f. = 1 p = 2.430496e-05

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

chi^2 = 16.16162 d.f. = 1 p = 5.816079e-05

실험설계 교차분석

- #05.상대위험율 계산
- $0.309/0.691$
- detach(precroData)

Console C:/Users/leecho/Desktop/RStat/Ch10/ ↗

```
> #05.상대위험율 계산  
> 0.309/0.691  
[1] 0.447178  
> |
```

사후설계 교차분석

■ 문제의 정의

- K병원에서는 흡연을 많이 하는 사람일수록 폐암에 걸릴 확률이 높다는 것을 발표하였다. 과연 흡연이 폐암과 연관이 있는지를 검증해 보자.
- 흡연자와 비흡연자를 대상으로 폐암발생여부를 파악하여 분석하고자 한다. 이때 국립보건소를 통해 다음과 같은 2차 자료를 구했다고 하자.
- 흡연을 많이 하면 폐암이 걸리는지를 검증해 보자(13.PostCRO.sav)

■ 가설

- 귀무가설(H_0): 흡연유무와 폐암유무와는 관계가 없다.

$$H_0 : \pi_1 = \pi_2$$

- 연구가설(H_1): 흡연유무와 폐암유무와는 관계가 있다.

$$H_1 : \pi_1 \neq \pi_2$$

사후설계 교차분석

- 오즈비(odds ratio)
 - 위험요인과 질병 발생간의 연관성을 1을 기준으로 나타낸 척도

		폐암유무	
		유(1)	무(2)
흡연유무	유(1)	73	18
	무(2)	141	196
계		214	214

$$odds_1 = \frac{73/214}{141/214} = 0.518$$

$$odds_2 = \frac{18/214}{196/214} = 0.092$$

$$OR = \frac{0.518}{0.092} = 5.63$$

* 흡연군의 폐암발생이 비흡연군에 비해 5.6배 많다



사후설계 교차분석

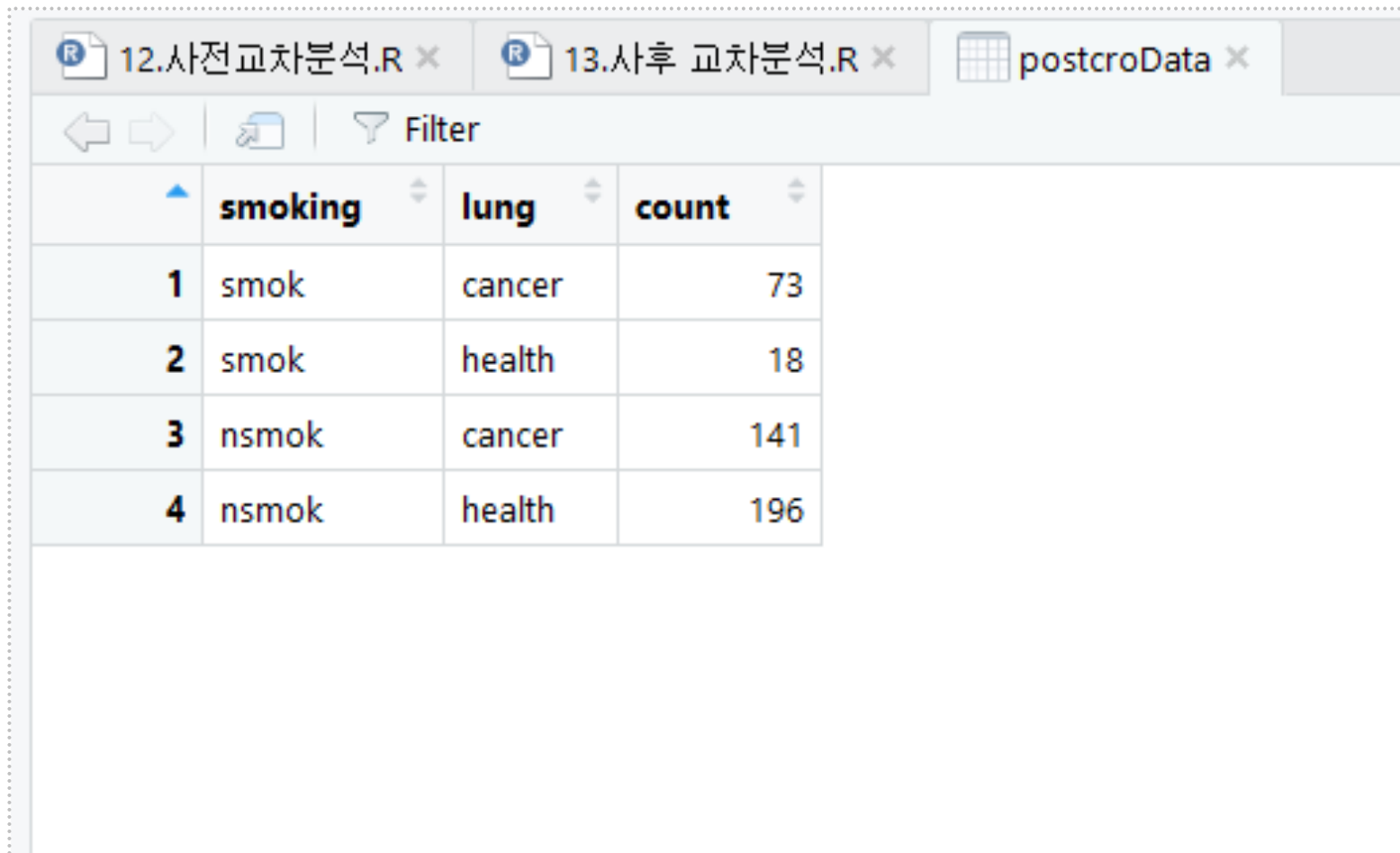
- ##### 사후 교차분석: 독립성 (chi-square) #####
- ### 표로 정리되어 있는 2차 데이터로 처리
- # 01.데이터 불러오기
- postcroData <- read.csv("./data/13.postCro.csv",
header=TRUE,
na.strings = ".")
- postcroData\$smoking <- factor(postcroData\$smoking,
levels=c(1,2),
labels=c("smok","nsmok"))
- postcroData\$lung <- factor(postcroData\$lung,
levels=c(1,2),
labels=c("cancer","health"))
- str(postcroData)



사후설계 교차분석

- `attach(postcroData)`
- #01.분할표 만들기
- `nTable <- xtabs(count ~ smoking + lung,
data=postcroData)`
- `addmargins(nTable)`
- `pTable <- prop.table(nTable)`
- `addmargins(pTable)`

사후설계 교차분석



The screenshot shows an RStudio window with three tabs: '12.사전교차분석.R', '13.사후 교차분석.R', and 'postcroData'. The '13.사후 교차분석.R' tab is active, displaying a data table with four columns: 'smoking', 'lung', and 'count'. The table contains four rows of data, numbered 1 to 4. The 'smoking' column has values 'smok' and 'nsmok', and the 'lung' column has values 'cancer' and 'health'. The 'count' column shows the frequency for each combination: 73 for (smok, cancer), 18 for (smok, health), 141 for (nsmok, cancer), and 196 for (nsmok, health).

	smoking	lung	count
1	smok	cancer	73
2	smok	health	18
3	nsmok	cancer	141
4	nsmok	health	196

사후설계 교차분석

Console C:/Users/leecho/Desktop/RStat/Ch10/ ↗

```
> attach(postcroData)
> #01. 분할표 만들기
> nTable <- xtabs(count ~ smoking + lung, data=postcroData)
> addmargins(nTable)
```

```
      lung
smoking cancer health Sum
smok      73      18  91
nsmok    141     196 337
Sum       214     214 428
```

```
> pTable <- prop.table(nTable)
> addmargins(pTable)
```

```
      lung
smoking  cancer    health      Sum
smok  0.17056075 0.04205607 0.21261682
nsmok 0.32943925 0.45794393 0.78738318
Sum    0.50000000 0.50000000 1.00000000
```



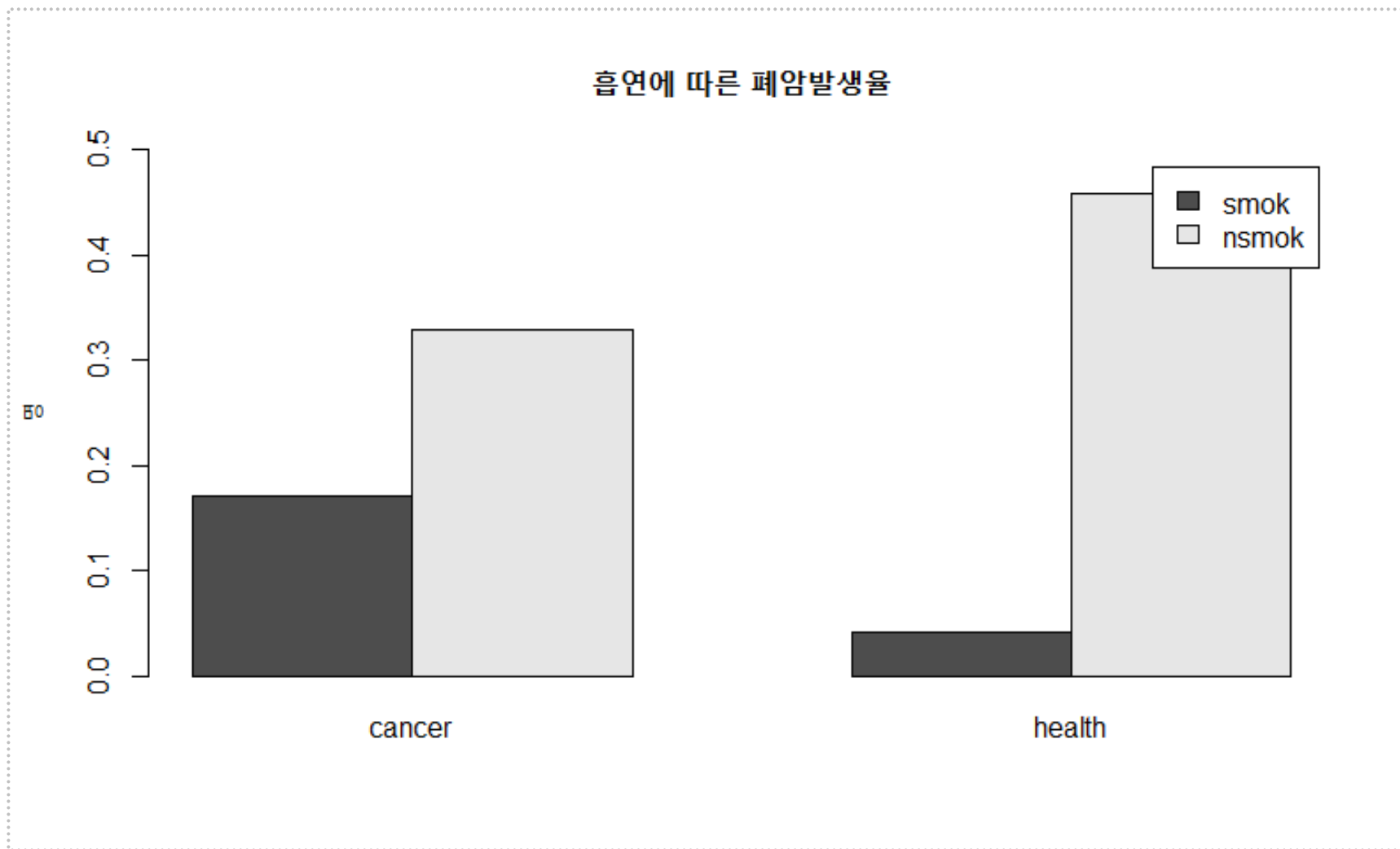

사후설계 교차분석

- #02.그래프 그리기
- `barplot(pTable, beside=TRUE,`
- `legend=TRUE, ylim=c(0, 0.5),`
- `ylab="명",`
- `main="흡연에 따른 폐암발생율")`

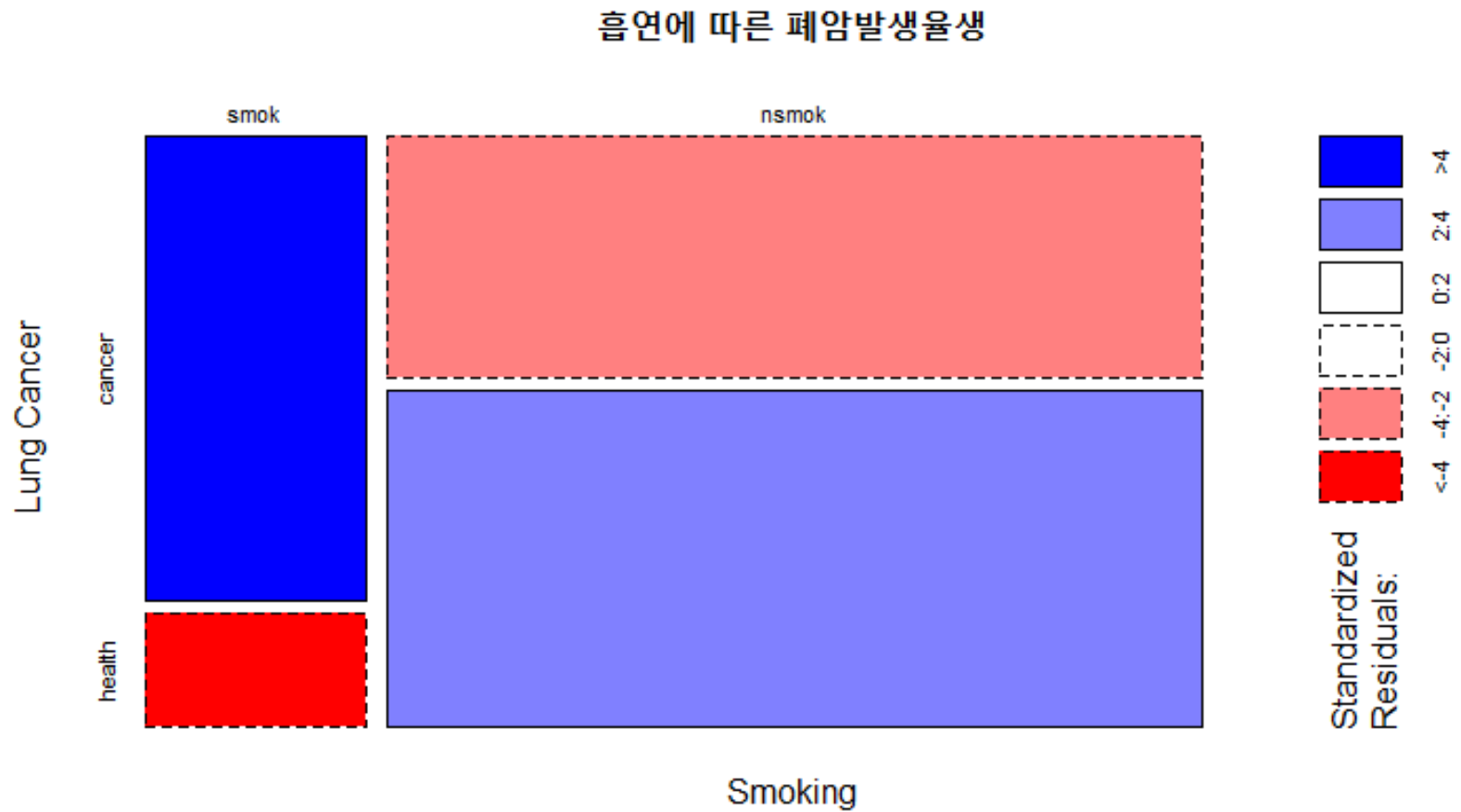
- # 모자이크 그래프
- `mosaicplot(nTable, shade=TRUE,`
- `xlab = "Smoking", ylab = "Lung Cancer",`
- `main = "흡연에 따른 폐암발생율생")`



사후설계 교차분석



사후설계 교차분석





사후설계 교차분석

- #03. 카이스케어 분석결과
- `chisq.test(nTable)`
- #04. 패키지 이용
- `install.packages("gmodels")`
- `library(gmodels)`
- `postcroResult <- CrossTable(nTable,`
- `expected=TRUE,`
- `chisq=TRUE)`
- `detach(postcroData)`

사후설계 교차분석

```
Console C:/Users/leecho/Desktop/RStat/Ch10/
> chisq.test(nTable)

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

data: nTable
X-squared = 40.697, df = 1, p-value = 1.778e-10

> #04.패키지이용
> install.packages("gmodels")
Error in install.packages : Updating loaded packages
> library(gmodels)
> postcroResult <- CrossTable(nTable,
+                             expected=TRUE,
+                             chisq=TRUE)

Cell Contents
-----
              N
Expected N
Chi-square contribution
N / Row Total
N / Col Total
N / Table Total
```

사후설계 교차분석

Console C:/Users/leecho/Desktop/RStat/Ch10/ ↗

	168.500	168.500	
	4.488	4.488	
	0.418	0.582	0.787
	0.659	0.916	
	0.329	0.458	
Column Total	214	214	428
	0.500	0.500	

Statistics for All Table Factors

Pearson's Chi-squared test

chi^2 = 42.21802 d.f. = 1 p = 8.164427e-11

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

chi^2 = 40.69678 d.f. = 1 p = 1.777803e-10

> |