

# 202021524\_성원호\_국가고시 교차분석

성원호

3/29/2021

## 교차분석

##(대립)가설: 국가고시 유형 별로 남녀의 합격자 수가 다를 것이다 ## 필요한 데이터를 수집하여, 교차분석을 통해 검증 ## 교차분석 도표 작성

```
##(대립)가설 : 국가고시 유형 별로 남녀의 합격자 수가 다를 것이다.
#install.packages("gmodels")
library(gmodels)
#install.packages("sjPlot")
library(sjPlot)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
## filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

```
glass <- read.csv("한국보건의료인국가시험_안경사.csv",fileEncoding = "CP949", header = T)
head(glass)
```

```
##   연도   직종 회차  성별 연령대   응시지역 졸업여부 합격여부
## 1 2000  안경사   12   여     30 서울특별시   졸업     합격
## 2 2000  안경사   12   남     20 서울특별시   졸업     불합격
## 3 2000  안경사   12   남     20 서울특별시   졸업     불합격
## 4 2000  안경사   12   여     40 서울특별시   졸업     합격
## 5 2000  안경사   12   남     20 서울특별시   졸업     합격
## 6 2000  안경사   12   남     20 서울특별시   졸업     합격
```

```
table_glass <- table(glass$합격여부, glass$성별)
table_glass
```

```
##
##           남      여
## 결시      2163    965
## 불합격    9025   5632
## 응시결격   71     49
## 합격     15481  13874
```

#table 함수를 사용하여 성별에 따른 합격 여부와 참여 유무의 합계를 구해줌

```
margin.table(table_glass)
```

```
## [1] 47260
```

```
margin.table(table_glass, 1)
```

```
##
## 결시 불합격 응시결격 합격
## 3128 14657 120 29355
```

```
margin.table(table_glass, 2)
```

```
##
## 남      여
## 26740 20520
```

#margin.table 함수를 사용하여 교차표의 주변합계를 계산해줌

#1 입력시 합격, 참여 컬럼을 기준으로 합계를 산출

#2 입력시 성별 기준으로 합계를 산출

```
round(prop.table(table_glass)*100, 2) #셀의 비율을 산출함
```

```
##
##           남      여
## 결시      4.58    2.04
## 불합격    19.10   11.92
## 응시결격   0.15    0.10
## 합격     32.76   29.36
```

```
round(prop.table(table_glass,2)*100, 2) #행의 비율을 산출함
```

```
##
##           남      여
## 결시      8.09    4.70
## 불합격    33.75   27.45
## 응시결격   0.27    0.24
## 합격     57.89   67.61
```

```
round(margin.table(prop.table(table_glass), 1)*100, 2) #행의 주변합계 비율 산출
```

```
##
##      결시      불합격 응시결격      합격
##      6.62      31.01      0.25      62.11
```

#round 함수를 사용해 소수점 두자리 수까지 반올림하며 백분을 산출을 위해 비율에 100을 곱해주었음.

```
chisq.test(table_glass, correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table_glass
## X-squared = 526.78, df = 3, p-value < 2.2e-16
```

#카이제곱 함수를 이용해 교차표에 나타난 성별에 따른 시험 합격 유무 정도를 통계적으로 검증

#그 결과  $\chi^2$ 값은 526.78이고 유의확률은  $2.2e-16$ 이 산출.

#p-value 값이 0.05보다 낮음을 확인할 수 있다.

#p-value가 0.05보다 낮으므로 귀무가설이 기각되고 대립가설인 성별에 따른 합격자 수가 다르다는 것이 받아들여진다.

```
CrossTable(glass$성별, glass$합격여부, digits=2,
            prop.c=TRUE, prop.r=FALSE, prop.t=FALSE, prop.chisq=FALSE,
            chisq=TRUE)
```

```
##
##
##      Cell Contents
## |-----|
## |                      N |
## |          N / Col Total |
## |-----|
##
##
## Total Observations in Table:  47260
##
##
##      | glass$합격여부
## glass$성별 |      결시 |      불합격 |      응시결격 |      합격 | Row Total |
## -----|-----|-----|-----|-----|-----|
##      남 |      2163 |      9025 |      71 |      15481 |      26740 |
##      |      0.69 |      0.62 |      0.59 |      0.53 |      |
## -----|-----|-----|-----|-----|
##      여 |      965 |      5632 |      49 |      13874 |      20520 |
##      |      0.31 |      0.38 |      0.41 |      0.47 |      |
## -----|-----|-----|-----|-----|
## Column Total |      3128 |      14657 |      120 |      29355 |      47260 |
##      |      0.07 |      0.31 |      0.00 |      0.62 |      |
## -----|-----|-----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----
## Chi^2 =  526.7846      d.f. =  3      p =  7.477533e-114
##
##
##
```

```
#gmodels 패키지에 내장된 CrossTable 함수로 편하게 교차표를 만들 수 있다.
#x^2 = 526.785, p-value = 0.000 이므로 0.05보다 낮음을 알 수 있음.
sjt.xtab(glass$성별, glass$합격여부, show.col.prc = T,
          show.exp = T, var.labels = c("성별", "합격여부"),
          value.labels = list(c("남자", "여자"),
                               c("합격", "불합격")),
          encoding="UTF-8")
```

```
## Warning: Length of `value.labels` does not match length of category values of
## `var.grp`.
```

		합격여부			Total
성별	결시	불합격	응시결격	합격	

	2163	9025	71	15481	26740
남자	1770	8293	68	16609	26740
	69.1 %	61.6 %	59.2 %	52.7 %	56.6 %
	965	5632	49	13874	20520
여자	1358	6364	52	12746	20520
	30.9 %	38.4 %	40.8 %	47.3 %	43.4 %
	3128	14657	120	29355	47260
<b>Total</b>	3128	14657	120	29355	47260
	100 %	100 %	100 %	100 %	100 %

$$\chi^2=526.785 \cdot df=3 \cdot \text{Cramer's } V=0.106 \cdot p=0.000$$

#sjPlot 패키지를 활용한 교차분석으로 viewer에 직접 출력해보았다.

```
sjt.xtab(glass$성별, glass$합격여부, show.col.prc = T,
         show.exp = T, var.labels = c("성별", "합격여부"),
         value.labels = list(c("남자", "여자"),
                             c("합격", "불합격")),
         file = "/Users/seong-wonho/Documents/glass_data.html")
```

```
## Warning: Length of `value.labels` does not match length of category values of
## `var.grp`.
```

성별	합격여부				Total
	결시	불합격	응시결격	합격	
	2163	9025	71	15481	26740
남자	1770	8293	68	16609	26740
	69.1 %	61.6 %	59.2 %	52.7 %	56.6 %
	965	5632	49	13874	20520
여자	1358	6364	52	12746	20520
	30.9 %	38.4 %	40.8 %	47.3 %	43.4 %
	3128	14657	120	29355	47260
<b>Total</b>	3128	14657	120	29355	47260
	100 %	100 %	100 %	100 %	100 %

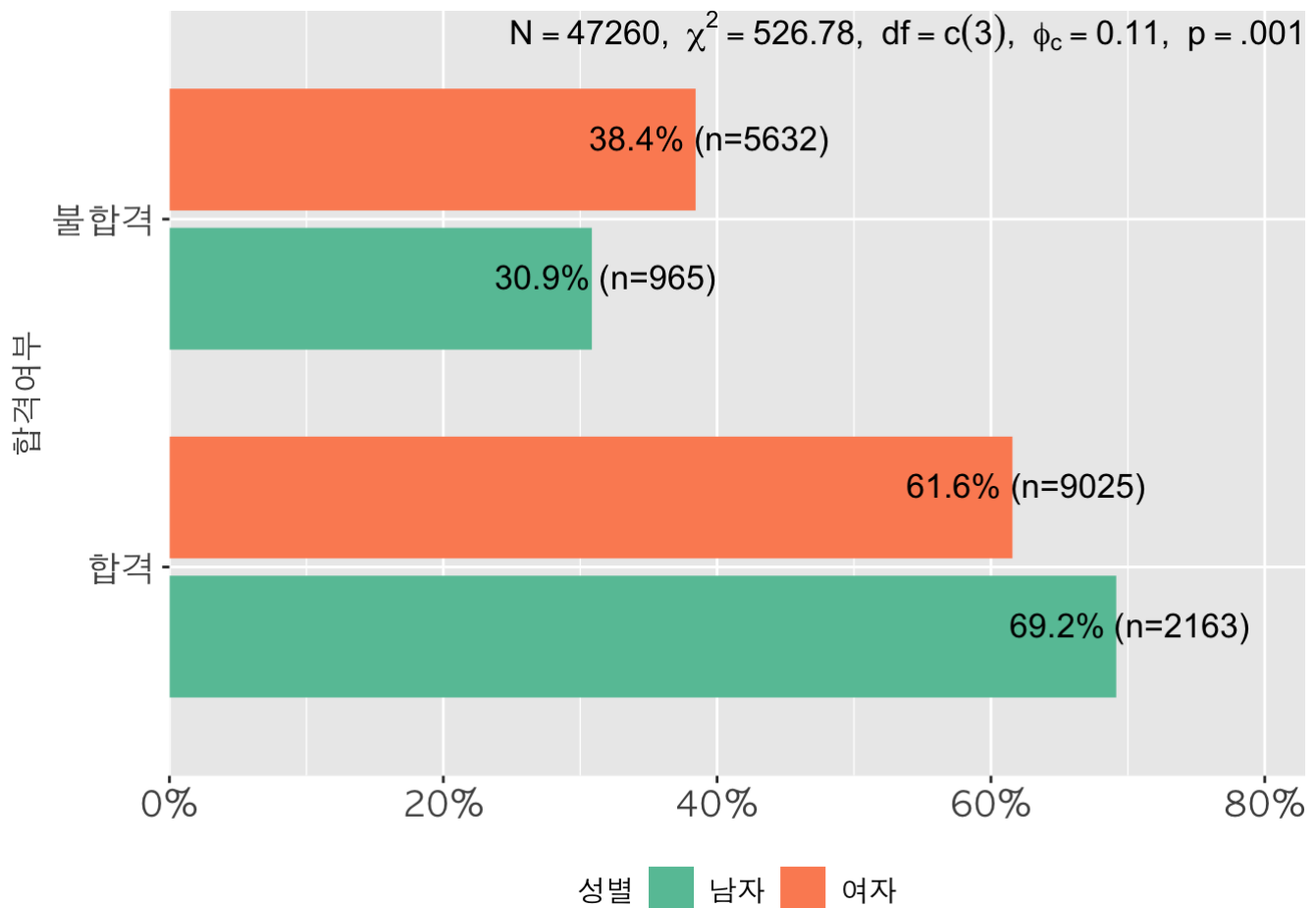
$$\chi^2=526.785 \cdot df=3 \cdot \text{Cramer's } V=0.106 \cdot p=0.000$$

```

browseURL("chisquare_table.html")
#이후 html 확장자로 저장
#  $\chi^2$  값은 526.785, p-value = 0.000 < 0.05라는 점을 알 수 있었음.

set_theme(geom.label.size = 4.5, axis.textsize = 1.1,
          legend.pos="bottom",
          theme.font = 'AppleGothic') #set_theme 함수를 활용해 도표 형성
plot_xtab(glass$성별, glass$합격여부, type="bar", #막대바 형태
          y.offset = 0.01,
          margin = "col", coord.flip = T, wrap.labels = 7,
          geom.colors = "Set2", show.summary = T, show.total = F,
          axis.titles = "합격여부",
          axis.labels = c("합격", "불합격"),
          legend.title = "성별",
          legend.labels = c("남자", "여자"))

```



```

physic <- read.csv("한국보건의료인국가시험원_응시.csv", fileEncoding = "CP949", header = T)

physic <- physic[-1, ]
#컬럼 내 오류가 있어 삭제 처리하였음.

table_physic <- table(physic$합격여부, physic$성별)
table_physic

```

```
##
##           남      여
## 결시      612    670
## 불합격    3772   6047
## 응시결격   34     35
## 합격     21762  39927
```

#table 함수를 사용하여 성별에 따른 합격 여부와 참여 유무의 합계를 구해줌

```
margin.table(table_physic)
```

```
## [1] 72859
```

```
margin.table(table_physic, 1)
```

```
##
##      결시   불합격  응시결격      합격
##      1282      9819          69    61689
```

```
margin.table(table_physic, 2)
```

```
##
##      남      여
## 26180 46679
```

#margin.table 함수를 사용하여 교차표의 주변합계를 계산해줌

#1 입력시 합격, 참여 컬럼을 기준으로 합계를 산출

#2 입력시 성별 기준으로 합계를 산출

```
round(prop.table(table_physic)*100, 2) #셀의 비율을 산출함
```

```
##
##           남      여
## 결시      0.84  0.92
## 불합격    5.18  8.30
## 응시결격  0.05  0.05
## 합격     29.87 54.80
```

```
round(prop.table(table_physic,2)*100, 2) #열의 비율을 산출함
```

```
##
##           남      여
## 결시      2.34  1.44
## 불합격   14.41 12.95
## 응시결격  0.13  0.07
## 합격     83.12 85.54
```

```
round(margin.table(prop.table(table_physic), 1)*100, 2) #행의 주변합계 비율 산출
```

```
##
##      결시      불합격 응시결격      합격
##      1.76      13.48      0.09      84.67
```

#round 함수를 사용해 소수점 두자리 수까지 반올림하며 백분을 산출을 위해 비율에 100을 곱해주었음.

```
chisq.test(table_physic, correct=FALSE)
```

```
##
## Pearson's Chi-squared test
##
## data:  table_physic
## X-squared = 120.76, df = 3, p-value < 2.2e-16
```

#카이제곱 함수를 이용해 교차표에 나타난 성별에 따른 시험 합격 유무 정도를 통계적으로 검증

#그 결과  $\chi^2$ 값은 120.76이고 유의확률은  $2.2e-16$ 이 산출.

#p-value 값이 0.05보다 낮음을 확인할 수 있다.

#p-value가 0.05보다 낮으므로 귀무가설이 기각되고 대립가설인 성별에 따른 합격자 수가 다르다는 것이 받아들여진다.

```
CrossTable(physic$성별, physic$합격여부, digits=2,
            prop.c=TRUE, prop.r=FALSE, prop.t=FALSE, prop.chisq=FALSE,
            chisq=TRUE)
```



```
##
##
##      Cell Contents
## |-----|
## |                      N |
## |          N / Col Total |
## |-----|
##
##
## Total Observations in Table:  72859
##
##
##      | physic$합격여부
## physic$성별 |      결시 |      불합격 |      응시결격 |      합격 | Row Total |
## -----|-----|-----|-----|-----|-----|
##      남 |      612 |      3772 |      34 |      21762 |      26180 |
##      |      0.48 |      0.38 |      0.49 |      0.35 |      |
## -----|-----|-----|-----|-----|
##      여 |      670 |      6047 |      35 |      39927 |      46679 |
##      |      0.52 |      0.62 |      0.51 |      0.65 |      |
## -----|-----|-----|-----|-----|
## Column Total |      1282 |      9819 |      69 |      61689 |      72859 |
##      |      0.02 |      0.13 |      0.00 |      0.85 |      |
## -----|-----|-----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----
## Chi^2 =  120.7556      d.f. =  3      p =  5.305087e-26
##
##
##
```

```
#gmodels 패키지에 내장된 CrossTable 함수로 편하게 교차표를 만들 수 있다.
#x^2 = 127.174, p-value = 5.056225e-25 이므로 0.05보다 낮음을 알 수 있음.
sjt.xtab(physic$성별, physic$합격여부, show.col.prc = T,
         show.exp = T, var.labels = c("성별", "합격여부"),
         value.labels = list(c("남자", "여자"),
                             c("합격", "불합격")),
         encoding="UTF-8")
```

```
## Warning: Length of `value.labels` does not match length of category values of
## `var.grp`.
```

		합격여부			Total
성별	결시	불합격	응시결격	합격	

	612	3772	34	21762	26180
남자	461	3528	25	22166	26180
	47.7 %	38.4 %	49.3 %	35.3 %	35.9 %
	670	6047	35	39927	46679
여자	821	6291	44	39523	46679
	52.3 %	61.6 %	50.7 %	64.7 %	64.1 %
	1282	9819	69	61689	72859
<b>Total</b>	1282	9819	69	61689	72859
	100 %	100 %	100 %	100 %	100 %

$$\chi^2=120.756 \cdot df=3 \cdot \text{Cramer's } V=0.041 \cdot p=0.000$$

#sjPlot 패키지를 활용한 교차분석으로 viewer에 직접 출력해보았다.

```
sjt.xtab(physic$성별, physic$합격여부, show.col.prc = T,
         show.exp = T, var.labels = c("성별", "합격여부"),
         value.labels = list(c("남자", "여자"),
                             c("합격", "불합격")),
         file = "/Users/seong-wonho/Documents/glass_data.html")
```

```
## Warning: Length of `value.labels` does not match length of category values of
## `var.grp`.
```

성별	합격여부				Total
	결시	불합격	응시결격	합격	
	612	3772	34	21762	26180
남자	461	3528	25	22166	26180
	47.7 %	38.4 %	49.3 %	35.3 %	35.9 %
	670	6047	35	39927	46679
여자	821	6291	44	39523	46679
	52.3 %	61.6 %	50.7 %	64.7 %	64.1 %
	1282	9819	69	61689	72859
<b>Total</b>	1282	9819	69	61689	72859
	100 %	100 %	100 %	100 %	100 %

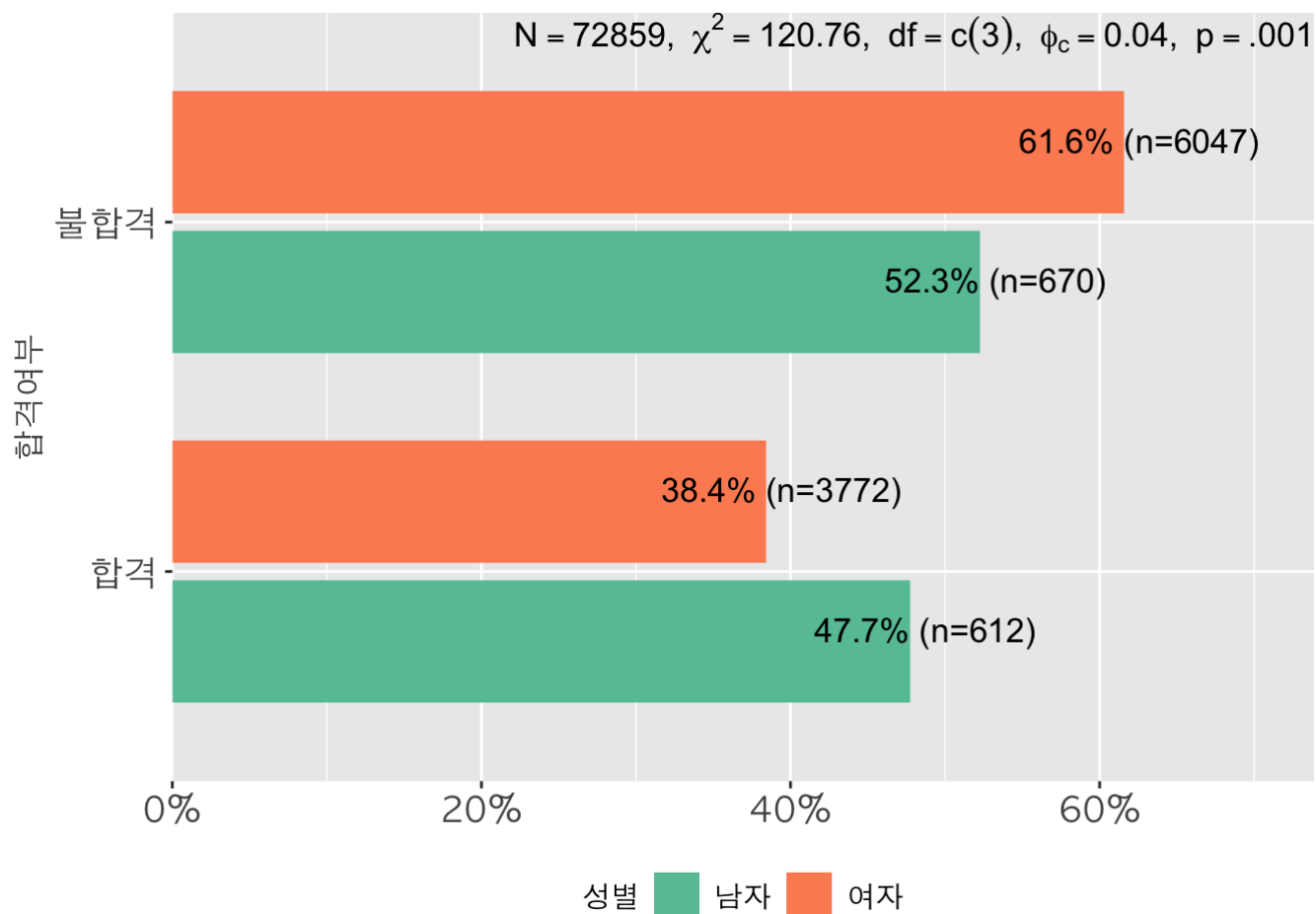
$$\chi^2=120.756 \cdot df=3 \cdot \text{Cramer's } V=0.041 \cdot p=0.000$$

```

browseURL("chisquare_table.html")
#이후 html 확장자로 저장
#  $\chi^2$  값은 127.174, p-value = 0.000 < 0.05라는 점을 알 수 있었음.

set_theme(geom.label.size = 4.5, axis.textsize = 1.1,
          legend.pos="bottom",
          theme.font = 'AppleGothic') #set_theme 함수를 활용해 도표 형성
plot_xtab(physic$성별, physic$합격여부, type="bar", #막대바 형태
          y.offset = 0.01,
          margin = "col", coord.flip = T, wrap.labels = 7,
          geom.colors = "Set2", show.summary = T, show.total = F,
          axis.titles = "합격여부",
          axis.labels = c("합격", "불합격"),
          legend.title = "성별",
          legend.labels = c("남자", "여자"))

```



```

test_table <- rbind(physic, glass)
head(test_table)

```

```
## 연도 직종 회차 성별 연령대 응시지역 졸업여부 합격여부
## 2 2000 물리치료사 27 여 20 서울특별시 졸업 합격
## 3 2000 물리치료사 27 남 20 서울특별시 졸업 결시
## 4 2000 물리치료사 27 여 20 서울특별시 졸업 불합격
## 5 2000 물리치료사 27 남 30 서울특별시 졸업 결시
## 6 2000 물리치료사 27 여 30 서울특별시 NULL 결시
## 7 2000 물리치료사 27 남 30 서울특별시 졸업 불합격
```

#유형별 합격자수 산출을 위해서는 2개 이상의 유형이 필요하므로 안경사와 물리치료사 데이터프레임을 결합함.

```
test_cal <- table(test_table$직종, test_table$합격여부)
test_cal
```

```
##
## 결시 불합격 응시결격 합격
## 물리치료사 1282 9819 69 61689
## 안경사 3128 14657 120 29355
```

#table 함수를 통해 직종에 따른 합격 여부의 합계를 구해줌

```
margin.table(test_cal)
```

```
## [1] 120119
```

```
margin.table(test_cal, 1)
```

```
##
## 물리치료사 안경사
## 72859 47260
```

```
margin.table(test_cal, 2)
```

```
##
## 결시 불합격 응시결격 합격
## 4410 24476 189 91044
```

#margin.table 함수를 사용하여 교차표의 주변합계를 계산해줌

#1 입력시 합격 컬럼을 기준으로 합계를 산출

#2 입력시 직종 기준으로 합계를 산출

```
round(prop.table(test_cal)*100, 2) #셀의 비율을 산출함
```

```
##
## 결시 불합격 응시결격 합격
## 물리치료사 1.07 8.17 0.06 51.36
## 안경사 2.60 12.20 0.10 24.44
```

```
round(prop.table(test_cal,2)*100, 2) #열의 비율을 산출함
```

```
##
##               결시 불합격 응시결격   합격
## 물리치료사 29.07  40.12    36.51 67.76
## 안경사      70.93  59.88    63.49 32.24
```

```
round(margin.table(prop.table(test_cal), 1)*100, 2) #행의 주변합계 비율 산출
```

```
##
## 물리치료사   안경사
##      60.66      39.34
```

#round 함수를 사용해 소수점 두자리 수까지 반올림하며 백분을 산출을 위해 비율에 100을 곱해주었음.

```
chisq.test(test_cal, correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  test_cal
## X-squared = 8140.3, df = 3, p-value < 2.2e-16
```

#카이제곱 함수를 이용해 교차표에 나타난 직종에 따른 시험 합격 유무 정도를 통계적으로 검증  
#그 결과  $\chi^2$ 값은 8140.3이고 유의확률은  $2.2e-16$ 이 산출.  
#p-value 값이 0.05보다 낮음을 확인할 수 있다.

```
CrossTable(test_table$직종, test_table$합격여부, digits=2,
            prop.c=TRUE, prop.r=FALSE, prop.t=FALSE, prop.chisq=FALSE,
            chisq=TRUE)
```

```
##
##
##      Cell Contents
## |-----|
## |                      N |
## |          N / Col Total |
## |-----|
##
##
## Total Observations in Table:  120119
##
##
##      | test_table$합격여부
## test_table$직종 |      결시 |      불합격 |      응시결격 |      합격 | Row Total |
## -----|-----|-----|-----|-----|-----|
##      물리치료사 |      1282 |      9819 |      69 |      61689 |      72859 |
##      |      0.29 |      0.40 |      0.37 |      0.68 |      |
## -----|-----|-----|-----|-----|-----|
##      안경사 |      3128 |      14657 |      120 |      29355 |      47260 |
##      |      0.71 |      0.60 |      0.63 |      0.32 |      |
## -----|-----|-----|-----|-----|-----|
##      Column Total |      4410 |      24476 |      189 |      91044 |      120119 |
##      |      0.04 |      0.20 |      0.00 |      0.76 |      |
## -----|-----|-----|-----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----
## Chi^2 =  8140.318      d.f. =  3      p =  0
##
##
##
```

#gmodels 패키지에 내장된 `CrossTable` 함수로 편하게 교차표를 만들 수 있다.  
 $\chi^2 = 8140.318$ ,  $p\text{-value} = 0.000$  이므로 0.05보다 낮음을 알 수 있음.  
 $p\text{-value}$ 가 0.05보다 낮으므로 귀무가설이 기각되고 대립가설인 국가고시 유형 별로 합격자 수가 다르다는 것이 받아들여진다.

```
sjt.xtab(test_table$직종, test_table$합격여부, show.col.prc = T,
         show.exp = T, var.labels = c("직종", "합격여부"),
         value.labels = list(c("물리치료사", "안경사"),
                             c("합격", "불합격")),
         encoding="UTF-8")
```

```
## Warning: Length of `value.labels` does not match length of category values of
## `var.grp`.
```

직종	합격여부				Total
	결시	불합격	응시결격	합격	
물리치료사	1282	9819	69	61689	72859
안경사	3128	14657	120	29355	47260
Column Total	4410	24476	189	91044	120119

	1282	9819	69	61689	72859
물리치료사	2675	14846	115	55223	72859
	29.1 %	40.1 %	36.5 %	67.8 %	60.7 %
	3128	14657	120	29355	47260
안경사	1735	9630	74	35821	47260
	70.9 %	59.9 %	63.5 %	32.2 %	39.3 %
	4410	24476	189	91044	120119
<b>Total</b>	4410	24476	189	91044	120119
	100 %	100 %	100 %	100 %	100 %

$$\chi^2=8140.318 \cdot df=3 \cdot \text{Cramer's } V=0.260 \cdot p=0.000$$

#sjPlot 패키지를 활용한 교차분석으로 viewer에 직접 출력해보았다.

```
sjt.xtab(test_table$직종, test_table$합격여부, show.col.prc = T,
  show.exp = T, var.labels = c("직종", "합격여부"),
  value.labels = list(c("물리치료사", "안경사"),
    c("합격", "불합격")),
  file = "/Users/seong-wonho/Documents/table_data.html")
```

```
## Warning: Length of `value.labels` does not match length of category values of
## `var.grp`.
```

직종	합격여부				Total
	결시	불합격	응시결격	합격	
	1282	9819	69	61689	72859
물리치료사	2675	14846	115	55223	72859
	29.1 %	40.1 %	36.5 %	67.8 %	60.7 %
	3128	14657	120	29355	47260
안경사	1735	9630	74	35821	47260
	70.9 %	59.9 %	63.5 %	32.2 %	39.3 %
	4410	24476	189	91044	120119
<b>Total</b>	4410	24476	189	91044	120119
	100 %	100 %	100 %	100 %	100 %

$$\chi^2=8140.318 \cdot df=3 \cdot \text{Cramer's } V=0.260 \cdot p=0.000$$

```

browseURL("chisquare_table.html")
#이후 html 확장자로 저장
#  $\chi^2$  값은 526.785, p-value = 0.000 < 0.05라는 점을 알 수 있었음.

set_theme(geom.label.size = 4.5, axis.textsize = 1.1,
          legend.pos="bottom",
          theme.font = 'AppleGothic') #set_theme 함수를 활용해 도표 형성
plot_xtab(test_table$직종, test_table$합격여부, type="bar", #막대바 형태
          y.offset = 0.01,
          margin = "col", coord.flip = T, wrap.labels = 7,
          geom.colors = "Set2", show.summary = T, show.total = F,
          axis.titles = "합격여부",
          axis.labels = c("합격", "불합격"),
          legend.title = "직종",
          legend.labels = c("물리치료사", "안경사"))

```

