

# 범주형자료분석방법론 HW1

이름: 김연주

학번: 2021250461

## 3.1

```
> #Table 3.1
> T3.1= matrix(c(54,10325,25,51790), nrow = 2, byrow = TRUE)
> rownames(T3.1)=c("No", "Yes")
> colnames(T3.1)=c("Fatal", "Nonfatal")
> T3.1
      Fatal Nonfatal
No       54    10325
Yes      25    51790
>
> library(epitools)
> epitools::oddsratio(T3.1,conf=0.95,correct=FALSE,method="wald")
$data
      Fatal Nonfatal Total
No       54    10325 10379
Yes      25    51790 51815
Total    79    62115 62194

$measure
              NA
odds ratio with 95% C.I. estimate      lower      upper
              No  1.00000              NA              NA
              Yes 10.83452  6.740538 17.41505

$p.value
      NA
two-sided midp.exact fisher.exact  chi.square
      No      NA      NA      NA
      Yes      0  2.66297e-24  6.754775e-35

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (wald) CI"
```

벨트를 하지 않았을 때 fatal injury일 오즈가 벨트를 착용했을 때 오즈의 10.83452배이다. 안전벨트를 하지 않았을 때의 부상이 안전벨트를 했을 때의 부상보다 치명적일 가능성이 높다.

95% CI가 (6.741,17.415)으로 1을 포함하지 않고 p-value가 0.05보다 작으므로 "오즈비가 1이다. 즉, 벨트 착용 여부와 부상 정도는 독립이다"라는 귀무가설이 기각된다.

## 3.2

```
> #Table 3.2
> T3.2=matrix(c(9,23,28,8,39,48,27,88,89,8,49,19,47,179,104,236,706,293),nrow=3)
> Degree=c("Less","HS","College")
> God=c("1","2","3","4","5","6")
> dimnames(T3.2)=list('Degree'= Degree, 'God'=God)
> T3.2
```

	God					
Degree	1	2	3	4	5	6
Less	9	8	27	8	47	236
HS	23	39	88	49	179	706
College	28	48	89	19	104	293

```
>
> chisq.test(T3.2)$expected
```

	God					
Degree	1	2	3	4	5	6
Less	10.05	15.9125	34.170	12.730	55.275	206.8625
HS	32.52	51.4900	110.568	41.192	178.860	669.3700
College	17.43	27.5975	59.262	22.078	95.865	358.7675

```
> chisq.test(T3.2)
```

Pearson's Chi-squared test

data: T3.2  
X-squared = 76.148, df = 10, p-value = 2.843e-12

```
> library(MASS)
> lrt=loglm(~ Degree + God, data = T3.2)
> summary(lrt)
```

Formula:  
~Degree + God  
attr("variables")  
list(Degree, God)  
attr("factors")  
Degree God  
Degree 1 0  
God 0 1  
attr("term.labels")  
[1] "Degree" "God"  
attr("order")  
[1] 1 1  
attr("intercept")  
[1] 1  
attr("response")  
[1] 0  
attr("Environment")  
<environment: R\_GlobalEnv>

Statistics:

	X^2	df	P(> X^2)
Likelihood Ratio	73.18791	10	1.070566e-11
Pearson	76.14833	10	2.842615e-12

```
>
```

- 1) 모든 cell의 기대빈도가 5 이상이므로 카이제곱 검정 실행
- 2) 카이제곱 검정 결과 p-value가 0.05 미만이므로 'Highest degree와 Belief in gods는 독립이다'는 귀무가설을 기각한다. 따라서 둘은 연관성을 지닌다.
- 3) LRT 결과 p-value가 0.05 미만이므로 동일한 귀무가설을 기각한다.

### 3.7

```
> #Table 3.7
> T3.7=matrix(c(13,23,14,29,59,67,15,47,54),nrow=3)
> PI=c("Liberal","Moderate","Conservative")
> Happiness=c("Not","Pretty","Very")
> dimnames(T3.7)=list('PI'= PI, 'Happiness'=Happiness)
> T3.7
```

PI	Happiness		
	Not	Pretty	Very
Liberal	13	29	15
Moderate	23	59	47
Conservative	14	67	54

```
>
> library(vcdExtra)
> CMHtest(T3.7,rscores = c(1,2,3),cscores=c(1,2,3))
Cochran-Mantel-Haenszel Statistics for PI by Happiness
```

	AltHypothesis	Chisq	Df	Prob
cor	Nonzero correlation	5.8517	1	0.015562
rmeans	Row mean scores differ	5.9139	2	0.051977
cmeans	Col mean scores differ	6.4629	2	0.039501
general	General association	7.0460	4	0.133475

```
> 1-pchisq(5.8517,df=1)
[1] 0.01556201
>
> Gkgamma(T3.7)
gamma      : 0.185
std. error : 0.078
CI         : 0.032 0.338
> z.value=0.185/0.078
> (1-pnorm(z.value))*2
[1] 0.01770192
>
```

- 1) Linear trend alternative to independence 이용
- 2) 귀무가설이 ' $\rho = 0$ ', 대립가설이 ' $\rho \neq 0$ '이라고 한다면, 검정통계량  $M^2$ 이 5.8517 이고 p-value가 0.015562으로 0.05보다 작다. 그러므로 귀무가설을 기각하고 PI와 Happiness 사이의 linear trend가 존재함을 알 수 있다.
- 3) Monotone trend 검정: gamma 검정통계량이 0.185, 양측검정 p-value가 0.0177로 귀무가설이 기각된다.

### 3.9

```
> #Table 3.9
> T3.9=matrix(c(3,1,1,3),nrow=2)
> Poured=c("Milk","Tea")
> Guess=c("Milk","Tea")
> dimnames(T3.9)=list('Poured'= Poured, 'Guess'=Guess)
> T3.9
      Guess
Poured Milk Tea
Milk      3   1
Tea       1   3
>
> chisq.test(T3.9)$expected
      Guess
Poured Milk Tea
Milk      2   2
Tea       2   2
경고메시지(들):
chisq.test(T3.9)에서: 카이제곱 approximation은 정확하지 않을수도 있습니다
> fisher.test(T3.9)

Fisher's Exact Test for Count Data

data:  T3.9
p-value = 0.4857
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.2117329 621.9337505
sample estimates:
odds ratio
 6.408309

> fisher.test(T3.9,alternative="greater")

Fisher's Exact Test for Count Data

data:  T3.9
p-value = 0.2429
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 0.3135693      Inf
sample estimates:
odds ratio
 6.408309
```

- 1) 모든 cell의 기대빈도가 5미만으로 피셔의 정확도 검정 실행
- 2) 양측검정과 단측검정에서 모두 p-value는 0.05 이상이므로 귀무가설 ' $\theta = 1$ '을 기각할 수 없다. 따라서, guess는 실제 순서와 무관하다.