범주형자료분석 HW1.

이름 : 김연주

학번 : 2021250461

학과 : 통계학과

1. R code:

```
> #1
> pi.hat=486/1500
> pi.hat-qnorm(0.995)*sqrt(pi.hat*(1-pi.hat)/1500)
[1] 0.2928744
> pi.hat+qnorm(0.995)*sqrt(pi.hat*(1-pi.hat)/1500)
[1] 0.3551256
>
```

$\therefore (0.29287,\ 0.35513)$

2. R code:

```
> #2
> pi.null=0.5
> z.test = (pi.hat-pi.null)/sqrt(pi.null*(1-pi.null)/1500)
> z.test
[1] -13.6329
> 2*pnorm(z.test)
[1] 2.552278e-42
>
```

a) $H_0: \pi_0 = \frac{1}{2}$ vs $H_a: \pi_0 \neq \frac{1}{2}$

b) Z test statistic = -13.6329

c) p-value = 2.2523e-42 < 0.05

d) reject H_0. 'Yes' 라고 답한 사람의 비율은 $\frac{1}{2}$이 아니다.

3. R code:

```
> #3
> pi1.hat=570/555453
> pi2.hat=433/8482
> prop.test(c(570,433), c(555453,8482), conf.level = 0.95, correct=FALSE)

	2-sample test for equality of proportions without continuity
	correction

data:  c(570, 433) out of c(555453, 8482)
X-squared = 11775, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.05470783 -0.04533835
sample estimates:
     prop 1      prop 2 
0.001026189 0.051049281 
```

$\therefore (-0.0547, -0.0453)$

4. R code:

```
> #4
> pic.hat=(570+433)/(555453+8482)
> z.stat = (pi1.hat-pi2.hat)/sqrt(pic.hat*(1-pic.hat)*(1/555453+1/8482))
> z.stat
[1] -108.5124
> 2*pnorm(z.stat)
[1] 0
>
```

(a) $H_0: \pi_1 - \pi_2 = 0$, $H_a: \pi_1 - \pi_2 \neq 0$

(b) I test statistic $= -108.5124$

(c) p-value ≈ 0 < 0.05

(d) reject H_0. 안전벨트를 착용했을 때와 착용하지 않았을 때의 치명상 확률이 다르다.

5. R code:

```
> #5
> chisq.stat=z.stat^2
> chisq.stat
[1] 11774.94
> #or
> data = matrix(c(570,433,554883,8049),nrow=2,ncol=2)
> chisq.test(data,correct=F)

        Pearson's Chi-squared test

data:  data
X-squared = 11775, df = 1, p-value < 2.2e-16
```

(a) $H_0: \pi_1 - \pi_2 = 0$, $H_a: \pi_1 - \pi_2 \neq 0$

(b) χ^2 test statistic = 11774.94

(c) p-value < 2.2e-16

(d) reject H_0. 안전벨트를 착용했을 때와 착용하지 않았을 때의 치명상 확률이 다르다.

6.

	fatal	nonfatal	
yes	570	554883	555453
no	433	8049	8482

$$\text{relative risk} = \frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{570/554883}{433/8482} = 0.0201$$

$$\text{odds ratio} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{570 \times 8049}{433 \times 554883} = 0.0191$$

$\therefore$ r.r. = 0.0201

o.r. = 0.0191

7. R code:

```
> #7
> data2 = matrix(c(802,34,53,494),nrow=2,ncol=2)
> chisq.stat2=chisq.test(data2,correct=F)
> ls(chisq.stat2)
[1] "data.name" "expected"  "method"    "observed"  "p.value"
[6] "parameter" "residuals" "statistic" "stdres"
> lrt.stat=2*(sum(chisq.stat2$observed*log(chisq.stat2$observed/chisq.stat2$e
xpected)))
> lrt.stat
[1] 1206.728
> 1-pchisq(lrt.stat,df=1)
[1] 0
```

(a) $H_0: \theta = 1$ vs $H_a: \theta \neq 1$

(b) LRT test statistic = 1206.728

(c) p-value = 0

(d) Reject H_0. 오즈비는 1이 아니다. 두 변수는 독립이 아니다.

8. R code:

```
> #8
> sample.or=(802*494)/(34*53)
> log.odds=log(sample.or)
> SE=sqrt(1/802+1/494+1/34+1/53)
> log.lb=log.odds-qnorm(0.975)*SE
> log.ub=log.odds+qnorm(0.975)*SE
> c(exp(log.lb),exp(log.ub))
[1] 140.8909 343.0917
>
```

$\therefore$ (140.8909, 343.0917)