**2021학년도 2학기 중간과제물(온라인제출용)**

**교과목명 : 보건정보데이터분석**

**학 번 : 202135-367895**

**성 명 : 김태정**

**연 락 처 : 010-4172-4516**

o 과제유형 : (공통) 형

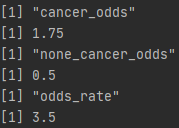
o 과 제 명 : 중간과제물

- 이하 과제 작성

**1번문제**

**(1)번문제**

cancer <- *c*(70, 30)  
none\_cancer <- *c*(40, 60)  
df <- *data.frame*(  
 *list*(cancer = cancer, none\_cancer = none\_cancer),  
 row.names = *c*('smoking', 'none\_smoking')  
)  
'cancer\_odds'  
(cancer\_odds <- (df[1, 1] / *sum*(df[, 1])) / (df[1, 2] / *sum*(df[, 2])))  
'none\_cancer\_odds'  
(none\_cancer\_odds <- (df[2, 1] / *sum*(df[, 1])) / (df[2, 2] / *sum*(df[, 2])))  
'odds\_rate'  
(odds\_rate <- cancer\_odds / none\_cancer\_odds)



폐암인 집단에서 흡연에 대한 오즈는 (70/(30+70))/(30/(30+70)) 이므로 1.75이다.

폐암이 아닌 집단에서의 흡연에 대한 오즈는 (40/(40+60))/(60/40+60) 이므로 0.5이다.

오즈비의 경우 폐암인 집단의 오즈/아닌집단의 오즈 이므로 1.75/0.5 이고 3.5이다.

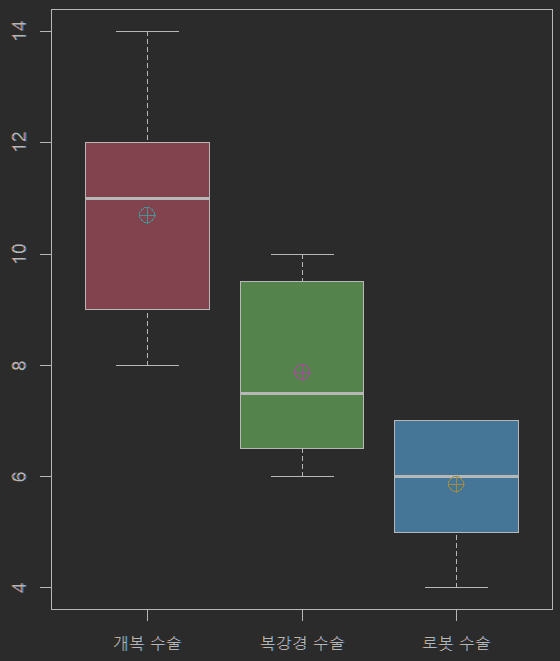
**(2)번문제**

상대위험률은 코호트 연구나 임상시험연구에서만 사용할 수 있다. 사례 대조군 연구(case-control study)는 사용할 수 없다. 그 이유는 상대위험률은 발생률을 구해야하는데 발생률이라는 건 전체 중에 얼마가 발생하는 비율을 의미한다. 전체 위험군 혹은 대조군이 있고 그 중에 위험률(발생률)을 알아야 하는데 사례 대조군 연구의 경우 위로 예를 들자면 흡연폐암값과 정상폐암값을 합치는게 의미가 없고 때문에 발생률을 알 수 가 없다. 발생률을 알려고 한다면 흡연 집단이 100명이고 그중에 폐암과 정상이 있는 형식이라면 가능할 것이다. 그렇기에 여기서는 오즈비(OR)을 사용하며 상대위험률(RR)을 사용할 수 없다.

**2번문제**

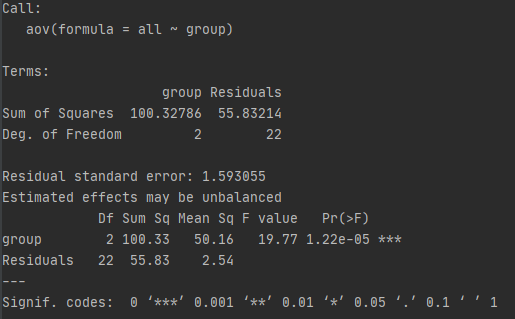
**(1)번문제**

open\_surgery <- *c*(12, 10, 14, 12, 11, 9, 9, 11, 11, 8)  
laparoscopic\_surgery <- *c*(9, 7, 8, 6, 6, 7, 10, 10)  
robot\_sergery <- *c*(5, 6, 7, 7, 7, 4, 5)  
*boxplot*(  
 open\_surgery,  
 laparoscopic\_surgery,  
 robot\_sergery,  
 col = 2:4,  
 names = *c*('개복 수술', '복강경 수술', '로봇 수술')  
)  
*points*(  
 *c*(  
 *mean*(open\_surgery),  
 *mean*(laparoscopic\_surgery),  
 *mean*(robot\_sergery)  
 ),  
 pch = 10,  
 cex = 2,  
 col = 5:7,  
)



**(2)번문제**

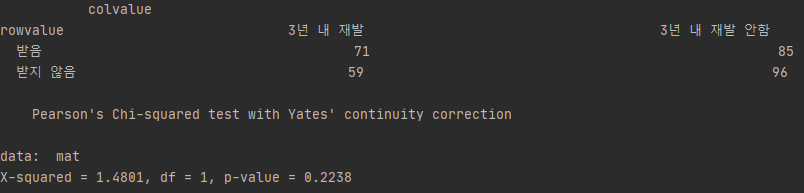
open\_surgery <- *c*(12, 10, 14, 12, 11, 9, 9, 11, 11, 8)  
laparoscopic\_surgery <- *c*(9, 7, 8, 6, 6, 7, 10, 10)  
robot\_sergery <- *c*(5, 6, 7, 7, 7, 4, 5)  
all <- *c*(open\_surgery, laparoscopic\_surgery, robot\_sergery)  
group <- *factor*(  
 *rep*(  
 *c*(  
 '개복 수술',  
 '복강경 수술',  
 '로봇 수술'  
 ),  
 *c*(  
 *length*(open\_surgery),  
 *length*(laparoscopic\_surgery),  
 *length*(robot\_sergery)  
 )  
 )  
)  
(result <- *aov*(all ~ group))  
*summary*(result)



유의 수준 5%보다 값이 작으므로 귀무가설은 기각된다. 즉 세 집단은 유의미한 차이가 있다고 본다.

**3번문제**

mat <- *matrix*(*c*(71, 85, 59, 96), nrow = 2, byrow = T)  
*dimnames*(mat) <- *list*(  
 rowvalue = *c*('받음', '받지 않음'),  
 colvalue = *c*('3년 내 재발', '3년 내 재발 안함')  
)  
mat  
*chisq.test*(mat)



유의 수준이 0.05보다 크므로 귀무가설을 채택한다.

**4번문제**

*library*(survival)  
  
kidney\_transplant\_patient <- *c*(  
 '66', '408', '425', '120+', '54',  
 '2150', '97+', '519+', '696', '873',  
 '1038', '633', '361+', '31', '1583'  
)  
time <- NULL  
status <- NULL  
for (atom in kidney\_transplant\_patient) {  
 if (ret <- *strsplit*(atom, '')[[1]][*nchar*(atom)] == '+') {  
 time <- *c*(  
 time,  
 *as.numeric*(*substr*(atom, 1, *nchar*(atom) - 1))  
 )  
 status <- *c*(  
 status,  
 0  
 )  
 }else {  
 time <- *c*(  
 time,  
 *as.numeric*(atom)  
 )  
 status <- *c*(  
 status,  
 1  
 )  
 }  
}  
df <- *data.frame*(time = time, status = status)  
result <- *survfit*(*Surv*(time, status) ~ 1, conf.type = 'log-log', data = df)  
*plot*(result, xlab = '시간', ylab = '생존함수', mark.time = T, col = *c*(2, 3, 3))  
*legend*(0.5, 0.2, *c*('누적한계추정값', '95% 신뢰구간'), lty = *c*(1, 2), col = 2:4)

