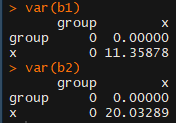
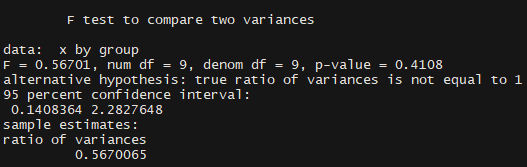
생명과학데이터분석 과제 #1

2014150137 통계학과

박정진

연습문제 2.1





서로 다른 비료로 자란 식물의 키에 대한 자료에서,

현재 사용하는 비료로 자란 식물의 평균과 분산, 표본분산을 각각 이라고 하며

새로운 비료로 자란 식물의 평균과 분산, 표본분산을 각각 라고 한다.

모분산 비교를 위하여 관심 있는 것은 두 모분산의 비, 이며,

이용하려는 검정 통계량은 이다.

검정하려는 가설은 vs. 이며,

F > 또는 F < 이면 귀무가설을 기각한다.

이때, 검정통계량 F = =0.56701 이며 P값은 0.4108이다.

이는 유의수준 로 했을 때, 귀무가설을 기각할만한 충분한 증거가 없으므로, 두 집단의 분산은 유의하게 다르지 않다고 볼 수 있다.

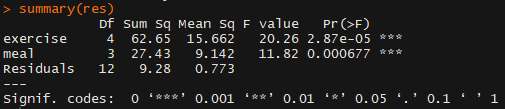
연습문제 2.7

(1)

식이요법과 운동방법에 따른 이원분산분석 모형은

이며, 이고 각각 독립인 오차항이다.

이때, 으로 제약한다.



여기서, 식이요법에 따라 체중 감량에 차이가 있는가?를 검정하기 위한 가설은

vs. 이며,

3번째 표의 meal을 보면, F value와 P value가 각각 11.82, 0.0007로, 유의수준 으로 했을 때, 귀무가설을 기각할만한 충분한 증거를 가지고 있으므로, 귀무가설을 유의하게 기각할 수 있다. 다시 말해, 식이요법에 따라 체중 감량에 차이가 있고 말할 수 있다.

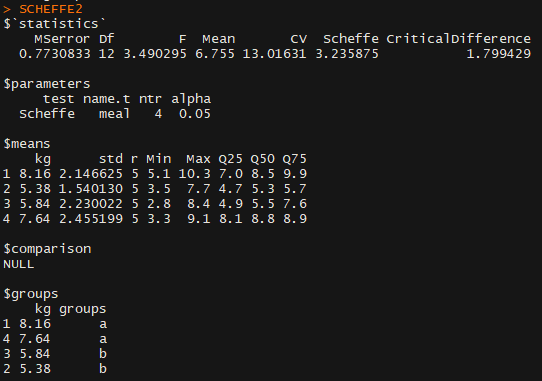
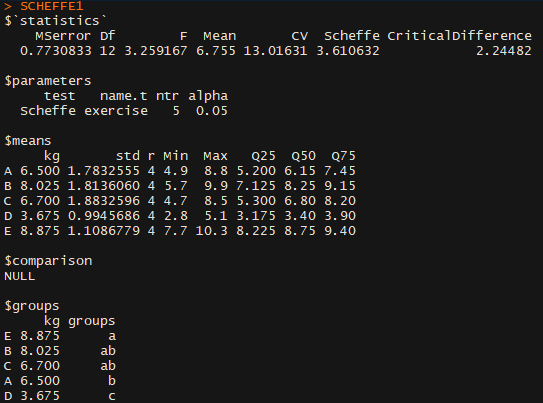
(2)

운동방법에 따라 체중 감량에 차이가 있는가?를 검정하기 위한 가설은

vs. 이며,

3번째 표의 exercise를 보면, F value와 P value가 각각 20.26, 0에 수렴하는 수로, 유의수준 으로 했을 때, 귀무가설을 기각할만한 충분한 증거를 가지고 있으므로, 귀무가설을 유의하게 기각할 수 있다. 다시 말해, 운동방법에 따라 체중 감량에 차이가 있고 말할 수 있다.

(3)

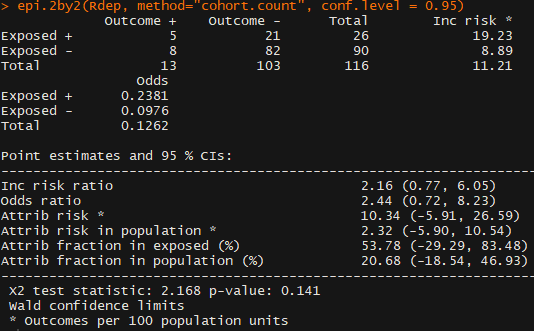


우선, 운동방법에 SCHEFFE의 방법을 실시한 결과 총 3개의 그룹으로 묶이게 되었으며, (운동방법E, 운동방법B) (운동방법 E, 운동방법C), (운동방법B, 운동방법C)사이에 유의미한 차이가 드러나지 않았다. 그리고 (운동방법B, 운동방법A), (운동방법C, 운동방법A)사이에서도 유의한 차이를 보이지 않았다. 하지만 그 외의 운동방법 간에는 유의한 차이를 보이고 있다.

식이요법에 SCHEFFE의 방법을 실시한 결과 총 2개의 그룹으로 묶이게 되었으며, (식이요법1, 식이요법4) (식이요법3, 식이요법2)끼리는 유의한 차이를 보이지 않았다. 하지만 그 외의 방법 간에는 유의한 차이를 보이고 있다.

연습문제 3.1





1. RR = =2.1635

OR = =2.4405

1. RR의 95% Confidence Interval = ( RR \* exp{}, RR \* exp{})

= (0.7734, 6.0516)

OR의 95% Confidence Interval = (OR \* exp{}, OR \* exp{})

= (0.7235, 8.2320)

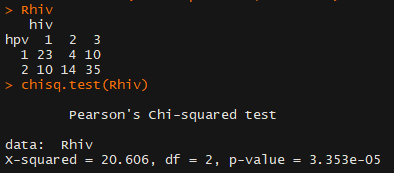
1. RR이 2.1635이기 때문에, 위험 요인이 있을 경우, 산후 우울증에 걸릴 위험은 위험 요인이 없을 경우 산후 우울증에 걸릴 위험의 2.1635배 라고 할 수 있다.

다만, 95%신뢰구간에 1이 포함되어있기 때문에 위험요인이 있는 그룹과 위험요인이 없는 그룹 간에 유의미한 차이가 있다고 보기는 어렵다.

OR이 2.4405이기 때문에, 산후 우울증에 걸린 집단 중에 위험요인이 있는 경우가 산후 우울증에 걸리지 않은 집단 중에 위험요인이 있는 경우보다 2.4405배라고 할 수 있다.

다만, 95%신뢰구간에 1이 포함되어있기 때문에 산후 우울증에 걸린 집단과 산후 우울증에 걸리지 않은 집단 간에 유의미한 차이가 있다고 말하기는 힘들다.

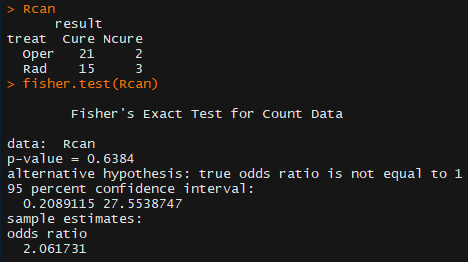
연습문제 3.4



셀의 기대 도수가 5 미만인 셀이 전체 셀의 20%를 초과하지 않기 때문에, 카이-스퀘어 검정을 사용할 수 있다.

이때, 검정 통계량은 = 20.6062이고, P-value는 0.0001보다 작기 때문에 유의수준 에서, HPV와 HIV의 감염 여부는 유의미한 관계성을 가진다고 볼 수 있다.

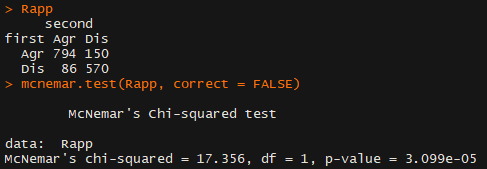
연습문제 3.5



기대도수가 5 미만인 셀이 전체의 20%를 초과하므로, Fisher 정확성 검정을 이용한다.

관측 된 값과 그것보다 극단적인 값이 나올 양측 검정의 p-value가 0.6384로, 유의수준 보다 크므로 귀무가설을 기각할 수 없고, 따라서 치료 결과와 방법 사이에는 연관성이 존재한다고 보기 어렵다고 할 수 있다.

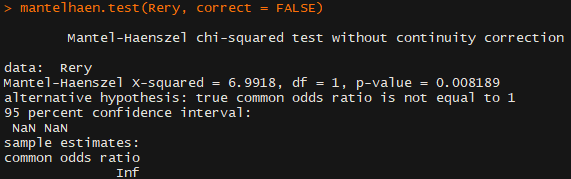
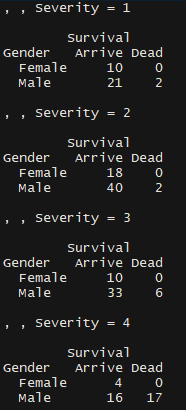
연습문제 3.7



동일 표본에 대하여 시간이 흐른 후 다시 조사를 한 것이기 때문에 맥니마 검정을 사용하였다.

여기서 나온 통계량이 17.356, P-value가 0.0001보다 작기 때문에, 유의수준 에서 귀무가설을 기각할 수 있으며, 첫 번째 투표와 두 번째 투표의 찬성확률은 동일하다고 볼 수 없다.

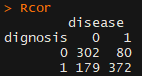
연습문제 3.8



질병의 심각도에 따라 층을 나누어 코크란-맨텔-헨젤방법을 이용하여 검정을 한 결과,

= 6.9918이며, p-value는 0.0082이다. 따라서 유의수준 에서 귀무가설을 기각할 수 있으며 남성 기증자의 수혈을 받은 경우와 여성 기증자의 수혈을 받은 경우의 사망률은 유의하게 다르다고 볼 수 있다.

연습문제 3.9



1. 민감도(Sensitivity) = ==0.62786

특이도(Specificity) = ==0.82301

1. 양성예측도 = P(D+ | T+) = = = 0.28272

음성예측도 = P(D- | T-) = = = 0.95216

<R 코드>

##SAS에서 입력한 데이터를 csv로 내보내어 사용하였습니다.

setwd("C:/Users/pkmon/Desktop/생과데1")

b <- read.csv("B.csv")

diet <- read.csv("diet.csv")

App <- read.csv("Approval.csv")

Can <- read.csv("CANCER.csv")

Coron <- read.csv("COR.csv")

Dep <- read.csv("Depress.csv")

Ery <- read.csv("ERY.csv")

Hiv <- read.csv("HIV.csv")

diet$meal <- as.factor(diet$meal)

**#2.1**

var(b1)

var(b2)

var.test(x ~ group, data=b, alternative = "two.sided")

**#2.7**

res <- aov(kg ~ exercise + meal, data=diet)

res

summary(res)

library(agricolae)

SCHEFFE1 <- scheffe.test(res, "exercise", group=TRUE, console=FALSE)

SCHEFFE2 <- scheffe.test(res, "meal", group=TRUE, console=FALSE)

SCHEFFE1

SCHEFFE2

**#3-1**

#(a)

library("epiR")

Dep$thr <- as.factor(Dep$thr)

Dep$dep <- as.factor(Dep$dep)

Rdep <- xtabs(obs ~ thr + dep, data=Dep)

Rdep

epi.2by2(Rdep, method="cohort.count", conf.level = 0.95)

**#3-4**

library("DescTools")

Rhiv <- xtabs(obs ~ hpv + hiv, data=Hiv)

Rhiv

chisq.test(Rhiv)

**#3-5**

Rcan <- xtabs(obs ~ treat + result, data=Can)

Rcan

fisher.test(Rcan)

**#3-7**

Rapp <- xtabs(count ~ first + second, data=App)

Rapp

mcnemar.test(Rapp, correct = FALSE)

**#3-8**

str(Ery)

Ery$Severity = factor(Ery$Severity, levels=unique(Ery$Severity))

Rery = xtabs(count ~ Gender + Survival + Severity, data=Ery)

Rery

mantelhaen.test(Rery, correct = FALSE)

Rery1 = xtabs(count ~ Gender + Survival, data=Ery)

chisq.test(Rery1)

**#3-9**

Rcor <- xtabs(count ~ dignosis + disease, data = Coron)

Rcor

Sensitivity <- Rcor[1,1]/sum(Rcor[,1])

Specificity <- Rcor[2,2]/sum(Rcor[,2])

PPT <- (Sensitivity\*0.1)/((Sensitivity\*0.1)+((1-Specificity)\*(1-0.1)))

PNT <- (Specificity\*0.9)/((Specificity\*0.9)+((1-Sensitivity)\*(0.1)))

result <- c(Sensitivity, Specificity, PPT, PNT)

names(result) <- c("Sensitivity", "Specificity", "PPT", "PNT")

result