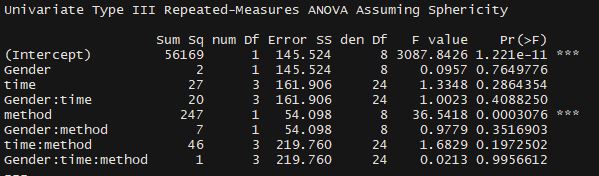
생명과학데이터분석 HW3

2014150137, 통계학과

박 정진

6-3

1. 반복 요인이 아니므로 구형성 검정의 필요성 X



Not

P-value가 0.7650으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 성별과 폐활량 지수의 증가 간에는 유의한 상관관계가 없다.

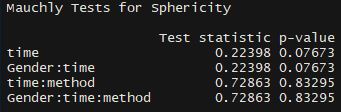
1. 운동 방법은 수준이 2개뿐으로, 별도의 구형성 자동 만족(일변량 분석)

Not

P-value가 0.0003으로 유의수준 =0.05하에서 귀무가설을 기각한다.

즉 운동 방법에 따라 폐활량 지수의 증가에 유의한 차이가 있다.

1. -1 구형성 검정



Not

P-value가 0.07673으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 시간은 구형성 가정을 만족한다.

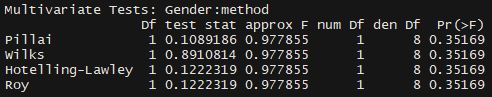
-2 일변량 분석

Not

P-value가 0.2864로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 시간에 따라 폐활량 지수가 유의하게 차이가 나지는 않는다.

1. –다변량



Not

P-value가 0.3517로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 두 요인 간에는 유의한 교호작용이 존재하지 않는다.

-일변량

Not

P-value가 0.3517로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 두 요인 간에는 유의한 교호작용이 존재하지 않는다.

두 방법 모두 동일하게 유의한 교호작용이 존재하지 않는 것으로 나타났다.

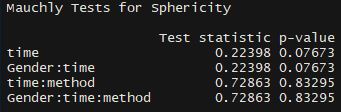
1. 시간이 구형성 가정을 만족하므로 일변량 분석을 이용하여 교호작용을 본다.

Not

P-value가 0.4088로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 성별과 시간 간에는 유의한 교호작용이 존재하지 않는다.

1. – 구형성 검정



Not

P-value가 0.8329으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 시간과 운동방법의 교호작용은 구형성 가정을 만족한다.

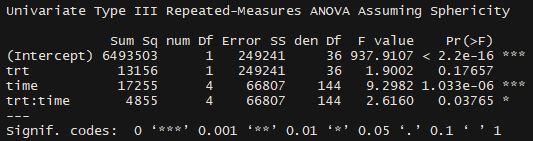
* 일변량 검정

Not

P-value가 0.1973으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉 운동 방법과 시간 간에는 유의한 교호작용이 존재하지 않는다.

6-6



1. 약의 종류는 반복 요인이 아니므로 구형성 검정 X

* 일변량 분석

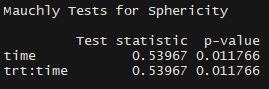
Not

P-value가 0.1766으로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉, 약의 종류에 따라 운동거리에 유의한 차이는 없다.

1. – 구형성 검정

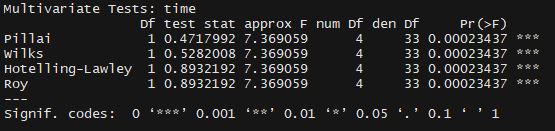
Not



P-value가 0.011766으로 유의수준 =0.05하에서 귀무가설을 기각한다.

즉 시간은 구형성 가정을 만족하지 못한다.

- 다변량 분석

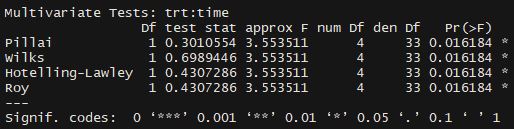


Not

P-value가 0.0002으로 유의수준 =0.05하에서 귀무가설을 기각한다.

즉, 시간에 따라 운동거리에 유의한 차이가 있다.

1. – 다변량



Not

P-value가 0.0162으로 유의수준 =0.05하에서 귀무가설을 기각한다.

즉, 시간에 따라 약의 종류 사이에 유의한 교호작용이 있다.

-일변량

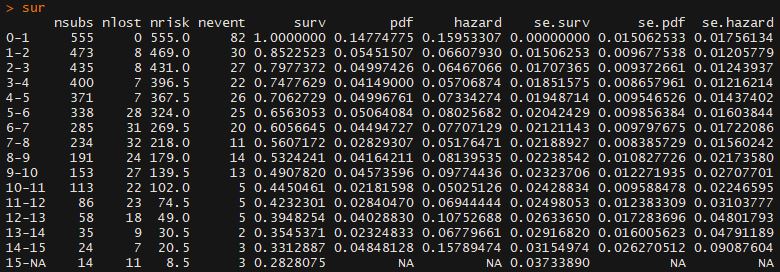
Not

P-value가 0.0376으로 유의수준 =0.05하에서 귀무가설을 기각한다.

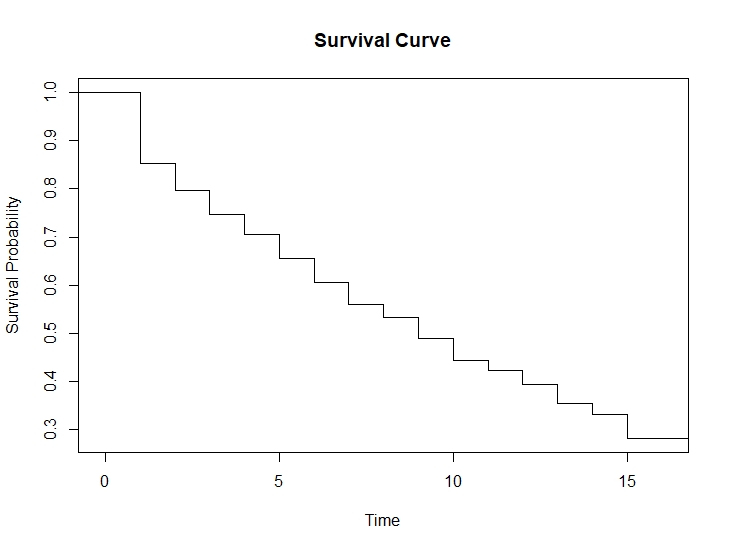
즉, 시간에 따라 약의 종류 사이에 유의한 교호작용이 있다.

두 분석 방법 모두에서 유의한 교호작용이 있음을 알 수 있다.

7-2

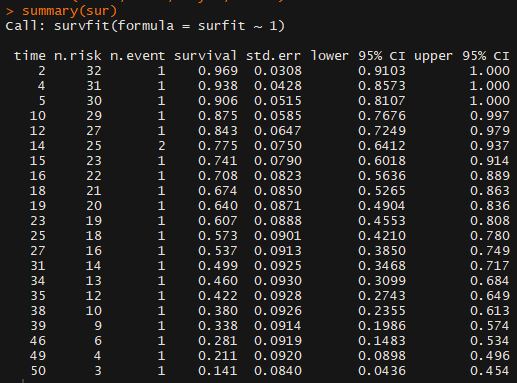


<추정된 각 구간별 유효인원수, 사망률, 생존율, 생존함수>

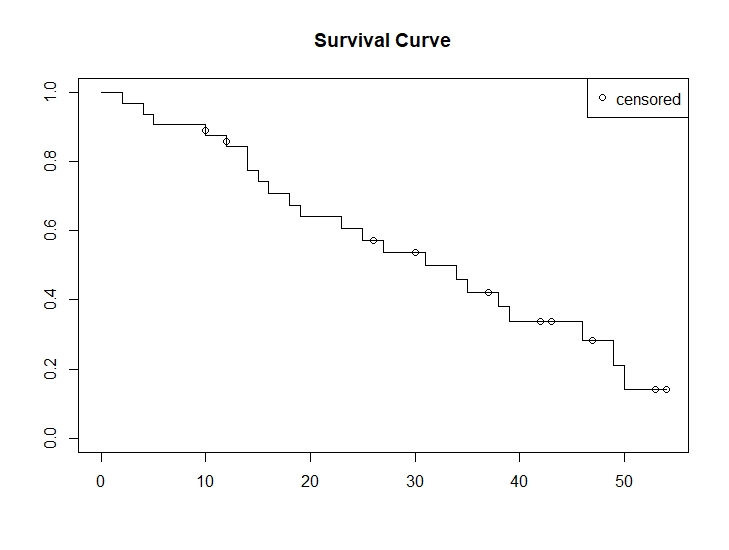


<생존 함수 그래프>

7-3



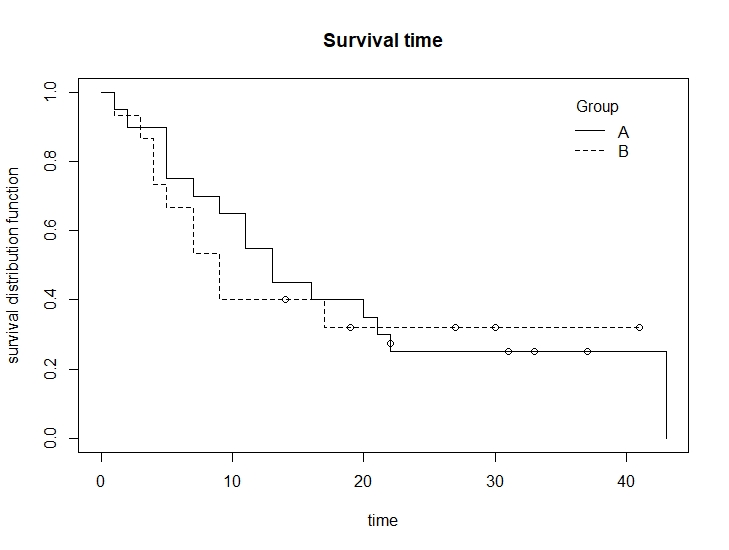
각 시점에서의 분산은 Std.err 행을 제곱함으로써 얻을 수 있다.



<생존함수의 그래프>

Time=31일 때, survival이 0.499로 대략 50%가 생존하는 것을 알 수 있다.

7-4



<생존함수 그래프>

생존 함수 그래프가 교차하므로, 로그-순위 검정법을 사용하기 적합하지는 않음을 알 수 있지만, 문제에서 요구한대로 검정을 수행한다.

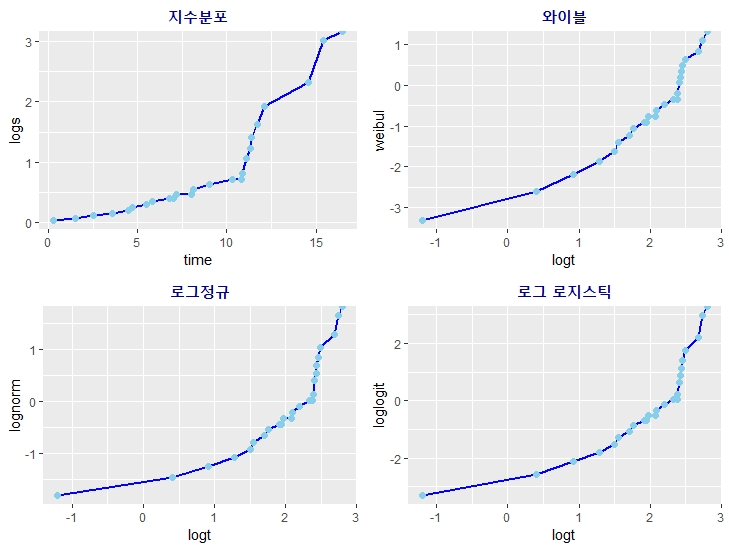
# 

Not

로그 순위 검정 통계량의 P-value가 0.8로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다.

즉, 두 그룹간 생존 함수는 유의미한 차이를 보이지 않는다.

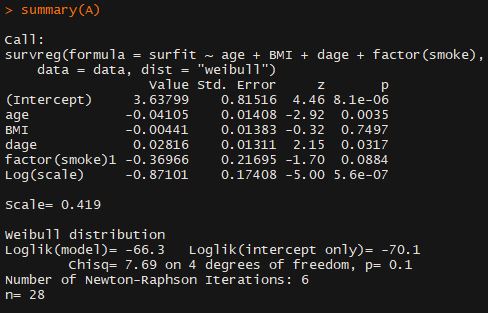
7-5



왼쪽 위부터 순서대로, 지수분포, 와이블 분포, 로그 정규 분포, 로그 로지스틱 분포의 그림.

지수분포를 제외하고는 대체로 직선 형태를 띠고 있다.

1. 와이블 분포를 이용하여 생존시간을 분석하였다.



분석하려는 모델은,

이다.

각 변수들에 대하여,

Not

를 유의수준 =0.05하에서 검정한 결과,

BMI와 흡연 상태는 귀무가설을 기각하지 못하여, 유의한 변수가 아닌 것으로 생각되며,

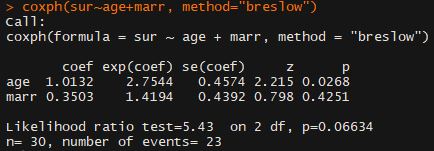
나이와 진단시 나이는 같은 유의수준에서 귀무가설을 기각하여 유의한 변수로 생각된다.

Not

의 우도비 검정 통계량의 경우,

-2(log L – log L0) = -2(-70.1+66.3)=7.69 이고 이는, 를 따른다. =9.84 이므로 유의수준 =0.05에서 귀무가설을 기각하지 못한다. 다만, P-value = 0.1로 유의수준과 큰 차이가 나지는 않는다.

7-6



* 나이

Not

카이스퀘어 검정 통계량의 P-value가 0.0268으로 유의수준 =0.05하에서 귀무가설을 기각한다. 즉, 나이에 대한 회귀 계수 1.01317은 유의하다고 할 수 있다.

* 골수 응고여부

Not

카이스퀘어 검정 통계량의 P-value가 0.4252로 유의수준 =0.05하에서 귀무가설을 기각하지 못한다. 즉, 골수 응고여부에 대한 회귀 계수 0.35025는 유의하지 않다고 할 수 있다.

다른 변인이 동일할 때, 나이가 50세 이상인 경우, 50세 미만인 경우보다 위험률이 2.754배 높다.

다른 변인이 동일할 때, 골수 응고가 100%인 경우, 그렇지 않은 경우보다 위험률이 1.419배 높다.

<R code>

library(KMsurv)

library(survival)

#7-2

time <- c(0:15 , NA)

death <- c(82,30,27,22,26,25,20,11,14,13,5,5,5,2,3,3)

cens <- c(0,8,8,7,7,28,31,32,24,27,22,23,18,9,7,11)

sur <- lifetab(tis = time, ninit = 555 , nlost = cens, nevent = death)

sur

plot(stepfun(x=1:15, y = sur$surv) , pch="", main = "Survival Curve",

ylab = "Survival Probability", xlab="Time")

#7-3

data <- read.csv("cancer2.csv")

surfit <- Surv(data$time,data$censor==1)

sur <- survfit(surfit~1)

plot(sur , conf.int = F, mark=1 , main = "Survival Curve")

legend("topright" , legend="censored", pch=1)

summary(sur)

#7-4

data <- read.csv("Hod.csv")

time <- data$time

cens <- data$censor

group <- data$group

surfit <- Surv(time , cens==1)

sur <- survfit(surfit~group)

plot(sur, col=c("red", "blue"),mark=1 , main = "Survival time" , xlab="time" ,ylab="survival distribution function")

legend(35,1.0 , legend=c("A","B"), lty=1, col=c("red", "blue") , title="Group", bty="n")

survdiff(surfit~group)

#7-5

library("flexsurv")

data <- read.csv("dia.csv")

data

time <- data$time

cens <- data$censor

surfit <- Surv(time , cens==1)

sur <- survfit(surfit~1)

S <- sur$surv

time<-unique(sort(time))

data1 <- cbind(S=S, logs=-log(S), weibul=log(-log(S)), lognorm=qnorm(1-S), loglogit=log((1-S)/S), logt=log(time))

data1 <- as.data.frame(data1)

library(ggplot2)

library(gridExtra)

#지수

p1<-ggplot(data=data1, aes(x=logt, y=logs))+

geom\_line(size=1, col="blue")+

geom\_point(size=2, col="skyblue")+

ggtitle("지수분포")+

theme(plot.title = element\_text(family = "serif", face = "bold", hjust = 0.5, size = 15, color = "darkblue"))

#와이블

p2<-ggplot(data=data1, aes(x=logt, y=weibul))+

geom\_line(size=1, col="blue")+

geom\_point(size=2, col="skyblue")+

ggtitle("와이블")+

theme(plot.title = element\_text(family = "serif", face = "bold", hjust = 0.5, size = 15, color = "darkblue"))

#로그-정규분포

p3<-ggplot(data=data1, aes(x=logt, y=lognorm))+

geom\_line(size=1, col="blue")+

geom\_point(size=2, col="skyblue")+

ggtitle("로그정규")+

theme(plot.title = element\_text(family = "serif", face = "bold", hjust = 0.5, size = 15, color = "darkblue"))

#로그-로지스틱

p4<-ggplot(data=data1, aes(x=logt, y=loglogit))+

geom\_line(size=1, col="blue")+

geom\_point(size=2, col="skyblue")+

ggtitle("로그 로지스틱")+

theme(plot.title = element\_text(family = "serif", face = "bold", hjust = 0.5, size = 15, color = "darkblue"))

grid.arrange(p1,p2,p3,p4, nrow=2, ncol=2)

#와이블 분포

time <- data$time

cens <- data$censor

A<-survreg(Surv(time, censor) ~ age + BMI + dage + factor(smoke), data = data, dist = "weibull")

summary(A)

#7-6

data <- read.csv("Leukemia.csv")

time <- data$time

cens <- data$censor

age <- data$age

marr <- data$x

sur = Surv(time, cens==1)

coxph(sur~age+marr, method="breslow")