**ADA Homework 10**

**Wenxin Liang wl2455**

Consider the Mayo Clinic Lung Cancer Data in R package survival: data (lung) or data (cancer): including the variables

inst: Institution code

time: Survival time in days

status: censoring status 1=censored, 2=dead

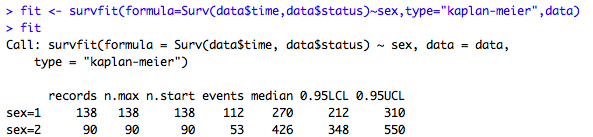
age: Age in years

sex: Male=1 Female=2, etc.

1. Estimate and plot the survival curves for time BY sex using the following methods:

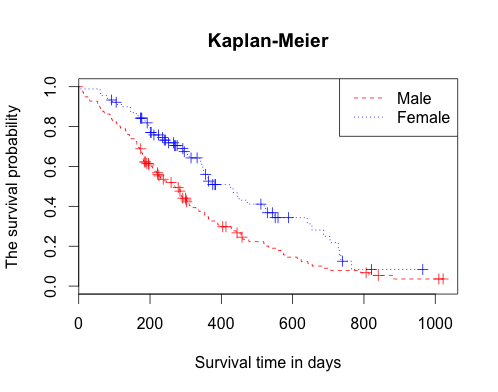
–  Kaplan-Meier

For Kaplan-Meier we obtain the survival curves by the following R code,



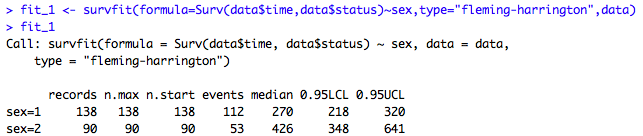
Since within the data we have our own status, we add status as another variable to influence sex to avoid using the default status.

The survival curves showed in graph as followed,



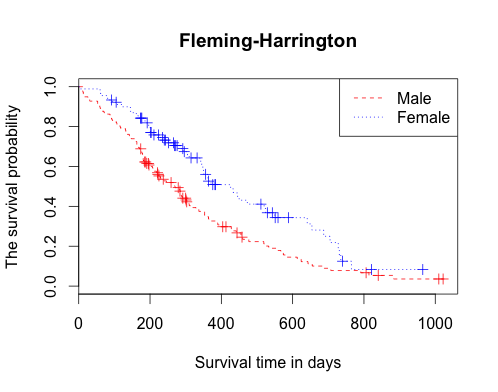
–  Fleming-Harrington

For Fleming-Harrington we obtain the survival curves by the following R code,



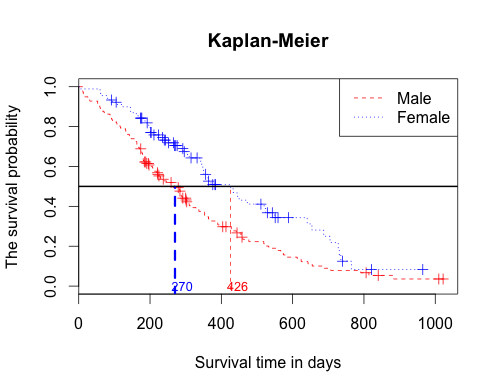
Since within the data we have our own status, we add status as another variable to influence sex to avoid using the default status.

The survival curves showed in graph as followed,

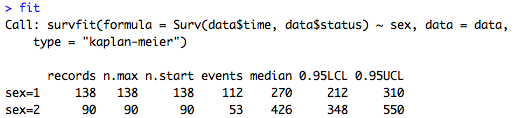


2. For each case in 1, estimate the median survival time, using the estimated survival curves.

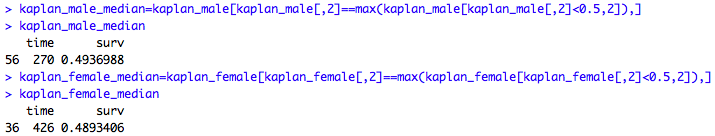
Based on the graph we get from Kaplan-Meier, we may conclude that the median survival time for male is 270 days and for female the survival time is 426 days.



We verify the medians we obtain by the survival curve indicated,

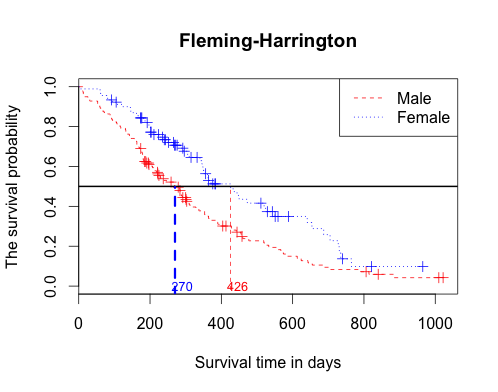


Moreover by the definition of median,

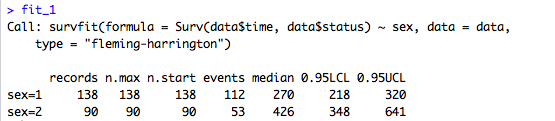


Therefore, we draw the conclusion as the median survival time for male is 270 days and for female the survival time is 426 days based on Kaplan-Meier method.

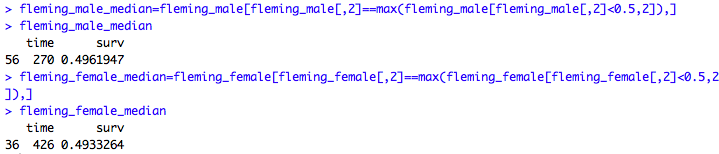
Based on the graph we get from Fleming-Harrington, we may conclude that the median survival time for male is 270 days and for female the survival time is 426 days.



We verify the medians we obtain by the survival curve indicated,



Moreover by the definition of median,



Therefore, we draw the conclusion as the median survival time for male is 270 days and for female the survival time is 426 days based on Fleming-Harrington.

Above all, we conclude that the median survival time for male is 270 days and for female the survival time is 426 days.

The following code,

#ADA Homework 10

library(survival)

data = lung

# Part I

# Kaplan-Meier

attach(data)

fit <- survfit(formula=Surv(data$time,data$status)~sex,type="kaplan-meier",data)

plot(fit,lty=2:4,main="Kaplan-Meier",xlab="Survival time in days",ylab="The survival probability",col=c("red","blue"))

legend("topright",c("Male","Female"),lty=2:4,col=c("red","blue"))

# Fleming-Harrington

fit\_1 <- survfit(formula=Surv(data$time,data$status)~sex,type="fleming-harrington",data)

fit\_1

plot(fit\_1,lty=2:4,main="Fleming-Harrington",xlab="Survival time in days",ylab="The survival probability",col=c("red","blue"))

legend("topright",c("Male","Female"),lty=2:4,col=c("red","blue"))

# Part II

# Kaplan-Meier

plot(fit,lty=2:4,main="Kaplan-Meier",xlab="Survival time in days",ylab="The survival probability",col=c("red","blue"))

legend("topright",c("Male","Female"),lty=2:4,col=c("red","blue"))

abline(h=0.5,lwd=2)

#calculate male

kaplan\_male=data.frame(time=summary(fit)$time[summary(fit)$strata=="sex=1"],surv=summary(fit)$surv[summary(fit)$strata=="sex=1"])

kaplan\_male

kaplan\_male\_median=kaplan\_male[kaplan\_male[,2]==max(kaplan\_male[kaplan\_male[,2]<0.5,2]),]

kaplan\_male\_median

lines(x=rep(kaplan\_male\_median[1],100),y=seq(-0.5,as.numeric(kaplan\_male\_median[2]),length=100),col="blue",lwd=3,lty=2)

text(x=kaplan\_male\_median[1]+20,y=0,labels=kaplan\_male\_median[1],cex=0.8,col="blue")

#calculate female

kaplan\_female=data.frame(time=summary(fit)$time[summary(fit)$strata=="sex=2"],surv=summary(fit)$surv[summary(fit)$strata=="sex=2"])

kaplan\_female

kaplan\_female\_median=kaplan\_female[kaplan\_female[,2]==max(kaplan\_female[kaplan\_female[,2]<0.5,2]),]

kaplan\_female\_median

lines(x=rep(kaplan\_female\_median[1],100),y=seq(-0.5,as.numeric(kaplan\_female\_median[2]),length=100),col="red",lwd=1,lty=2)

text(x=kaplan\_female\_median[1]+20,y=0,labels=kaplan\_female\_median[1],cex=0.8,col="red")

fit

# Fleming-Harrington

plot(fit\_1,lty=2:4,main="Fleming-Harrington",xlab="Survival time in days",ylab="The survival probability",col=c("red","blue"))

legend("topright",c("Male","Female"),lty=2:4,col=c("red","blue"))

abline(h=0.5,lwd=2)

#calculate male

fleming\_male=data.frame(time=summary(fit\_1)$time[summary(fit\_1)$strata=="sex=1"],surv=summary(fit\_1)$surv[summary(fit\_1)$strata=="sex=1"])

fleming\_male

fleming\_male\_median=fleming\_male[fleming\_male[,2]==max(fleming\_male[fleming\_male[,2]<0.5,2]),]

fleming\_male\_median

#abline male

lines(x=rep(fleming\_male\_median[1],100),y=seq(-0.5,as.numeric(fleming\_male\_median[2]),length=100),col="blue",lwd=3,lty=2)

text(x=fleming\_male\_median[1]+20,y=0,labels=fleming\_male\_median[1],cex=0.8,col="blue")

#calculate female

fleming\_female=data.frame(time=summary(fit\_1)$time[summary(fit\_1)$strata=="sex=2"],surv=summary(fit\_1)$surv[summary(fit\_1)$strata=="sex=2"])

fleming\_female

fleming\_female\_median=fleming\_female[fleming\_female[,2]==max(fleming\_female[fleming\_female[,2]<0.5,2]),]

fleming\_female\_median

#abline female

lines(x=rep(fleming\_female\_median[1],100),y=seq(-0.5,as.numeric(fleming\_female\_median[2]),length=100),col="red",lwd=1,lty=2)

text(x=fleming\_female\_median[1]+20,y=0,labels=fleming\_female\_median[1],cex=0.8,col="red")

fit\_1