############################################

# #

# Practical 2 #

# #

############################################

########################################

# Exercise 1

########################################

library(survival) # download the package of survival data analysis

help(lung) # help on lung

help(survfit)

# Kaplan-Meier estimation of the survival function

kmfit<-survfit(Surv(time,status)~ 1,data=lung,conf.type="plain",type='kaplan-meier')

print(kmfit)

summary(kmfit)

plot(kmfit)

plot(kmfit,mark.time=F,xscale=365.25,xlab="Time (in years)",ylab="Survival S(t)")

legend(1,0.8, c("Kaplan-Meier estimator", "95% pointwise CI"), lty=1:2)

# Harrington-Fleming estimation of the survival function

fhfit<-survfit(Surv(time,status)~1,data=lung,conf.type="plain",type='fh')

plot(kmfit,mark.time=F,xscale=365.25,xlab="Time (in years)",ylab="Survival S(t)")

lines(fhfit,lty=3,mark.time=F,xscale=365.25,col="red")

plot(kmfit$time,kmfit$surv-fhfit$surv)

# Nelson-Aalen estimation of the cumulative hazard function

naH =-log(fhfit$surv)

time= fhfit$time

plot(time,naH,type="s",ylab="Cumulative risk H(t)",xlab="Time (in months)")

##############################################

# Exercise 2: Freireich leukemia

##############################################

# 6-MP data

time.6mp <- c(6,6,6,6,7,9,10,10,11,13,16,17,19,20,22,23,25,32,32,34,35)

status.6mp <- c(1,1,1,0,1,0,1,0,0,1,1,0,0,0,1,1,0,0,0,0,0)

# Pacebo data

time.placebo <- c(1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23)

status.placebo <- rep(1,21)

# Gathered data

group <- c(rep(0,21),rep(1,21))

time <- c(time.placebo,time.6mp)

status <- c(status.placebo, status.6mp)

# Transformation of the data in "censored" data

help(Surv)

freireich <- Surv(time,status)

summary(survfit(freireich~group))

plot(survfit(freireich~group,conf.type="none"),lty=1:2,xlab="Survival time in months",ylab="Survival rate")

# Transformation of the data in "censored" data by group

freireich.placebo <- Surv(time.placebo,status.placebo)

summary(survfit(freireich.placebo)~1)

plot(survfit(freireich.placebo~1,conf.type="none"),lty=1:2,xlab="Survival time in months",ylab="Survival rate")

freireich.6mp <- Surv(time.6mp,status.6mp)

summary(survfit(freireich.6mp)~1)

plot(survfit(freireich.6mp~1,conf.type="none"),lty=1:2,xlab="Survival time in months",ylab="Survival rate")

##########################################

# Exercise 3: Hepatitis data

##########################################

# Data

time=c(1,1,1,1,4,5,7,8,10,10,12,16,16,16)

status=c(1,1,1,0,0,1,1,1,1,0,0,0,0,0)

# Kaplan-Meier estimation of the survival function

kmfit=survfit(Surv(time,status)~1,conf.type="none")

summary(kmfit)

plot(kmfit,mark.time=F,col="blue",xlab="Time (in years)",ylab="Survival S(t)")

# Harrington-Fleming estimation of the survival function

fhfit=survfit(Surv(time,status)~1,conf.type="none",type="fh")

summary(fhfit)

lines(fhfit,lty=2,mark.time=F,col="red")

legend(8,0.9, c("Kaplan-Meier", "Fleming-Harrington"), lty=1:2,col=c("blue","red"))

lambda.chap=sum(status)/sum(time)

# or adjustment of a parametric model

help(survreg)

# caution:

# survreg's scale = 1/(rweibull shape)

# survreg's intercept = log(rweibull scale) that is 1/scale with respect to ours

survreg(Surv(time,status)~1,dist="exponential")

# thus lambda.chap=exp(-intercept)

tt=seq(0,15,0.01)

expsurv=exp(-lambda\*tt)

lines(tt,expsurv,lty=4,col="magenta")

#########################################

# Exercise 4: creating data

#########################################

########################################

# Simulation of n realizations of Exp(1.1)

n=100

X=rexp(n,1.1)

# Empirical cdf

plot(ecdf(X))

# Kaplan-Meier estimation of the survival function

plot(survfit(Surv(X)~1,conf.type="none",type='kaplan-meier'))

# Theoretical survival function

x=seq(0,4,0.01)

S.theo=exp(-1.1\*x)

lines(x,S.theo,lty=2,col="red")

########################################

# Simulation of n censored data

C=rexp(n,1)

T=pmin(X,C)

delta=(X<=C)

nd=sum(delta)

# Whole observations (theoretical, Kaplan-Meier, MLE)

kmfit.T=survfit(Surv(T,delta)~1,type='kaplan-meier',conf.type="none")

plot(kmfit.T,mark.time=F)

lines(x,S.theo,lty=2,col="red")

lambda.mle=nd/sum(T)

S.mle=exp(-lambda.mle\*x)

lines(x,S.mle,lty=2,col="magenta")

# Uncensored data (Kaplan-Meier, MLE)

T.uncensored=T[delta==1]

kmfit.T.uncensored=survfit(Surv(T.uncensored)~1,type='kaplan-meier',conf.type="none")

x=seq(0,2.5,0.01)

lines(kmfit.T.uncensored,col="blue")

lambda.mle.uncensored=nd/sum(T.uncensored)

S.mle.uncensored=exp(-lambda.mle.uncensored\*x)

lines(x,S.mle.uncensored,lty=2,col="green")

########################################

# Same question with n=1000

########################################

# Comparison of the confidence intervals (plain, log, log-log)

x=seq(0,4,0.01)

par(mfrow=c(2,2))

kmfit.T.plain=survfit(Surv(T,delta)~1,type='kaplan-meier',conf.type="plain")

kmfit.T.log=survfit(Surv(T,delta)~1,type='kaplan-meier',conf.type="log")

kmfit.T.log2=survfit(Surv(T,delta)~1,type='kaplan-meier',conf.type="log-log")

plot(kmfit.T.plain,mark.time=F)

lines(x,S.theo,col="red")

title("CI plain")

plot(kmfit.T.log,mark.time=F)

lines(x,S.theo,col="red")

title("CI log")

plot(kmfit.T.log2,mark.time=F)

lines(x,S.theo,col="red")

title("CI log-log")

############################################

# #

# Practical 3 #

# #

############################################

###########################################

# Exercise 2: log-rank breast

###########################################

# 45 women 2 groups

breast=matrix(scan("breast.txt"),ncol=3,byrow=T)

breast=as.data.frame(list(duree=breast[,1],cens=breast[,2],gpe=breast[,3]))

# Kaplan-Meier estimation for both groups

kmbreast=survfit(Surv(breast$duree,breast$cens)~breast$gpe,type='kaplan-meier',conf.type="log-log")

summary(kmbreast)

plot(kmbreast,lty=3:2,xlab="Survival time",ylab="Survival")

# Logrank test

survdiff(Surv(breast$duree,breast$cens)~breast$gpe,rho=0)

# Harrington-Flemming test

survdiff(Surv(breast$duree,breast$cens)~breast$gpe,rho=1)

###################################################

# Practical 3 - Exercise 3: stratified logrank test

###################################################

# Data

time=c(8,8,13,18,23,52,63,63,70,76,180,195,210,220,365,632,700,852,1296,1296,1328,1460,1976,1990,2240)

status=c(rep(1,14),0,1,1,0,1,rep(0,6))

group=c(1,1,2,2,2,1,1,1,2,2,2,2,2,1,1,2,2,1,2,1,1,1,1,2,2) # treatment group

fctren=c(0,1,rep(0,6),rep(1,17)) # kidney function: 0 = abnormal,1= normal

# Logrank test to compare the survival function according to the kidney function

survdiff(Surv(time,status)~fctren) # S

# Plot of the survival functions according to the kifney function or the group

plot(survfit(Surv(time,status)~fctren))

plot(survfit(Surv(time,status)~group))

# Logrank test to compare the survival function of the two groups

survdiff(Surv(time,status)~group) # NS

plot(survfit(Surv(time,status)~group,subset=(fctren==1)))

plot(survfit(Surv(time,status)~group,subset=(fctren==0)))

# Logrank test to compare the survival function of the two groups stratifying on the kidney function

survdiff(Surv(time,status)~group,subset=(fctren==1)) # NS (limit)

survdiff(Surv(time,status)~group,subset=(fctren==0)) # NS

# Validation

survdiff(Surv(time,status)~group+ strata(fctren)) # becomes S !