Epidemiologic data analysis using R

Practicals 9

Faculty of Social Sciences, University of Tampere

–

Janne Pitkaniemi

09.04.2018

library(Epi)

Warning: package 'Epi' was built under R version 3.4.4

library(mstate)

Loading required package: survival

library(survival)  
library(lubridate)

Warning: package 'lubridate' was built under R version 3.4.3

Attaching package: 'lubridate'

The following object is masked from 'package:base':  
  
 date

# Survival analysis: Oral cancer patients

## Description of the data

File , that you may access from a url address to be given in the practical, contains data from 338 patients having an oral squamous cell carcinoma diagnosed and treated in one tertiary level oncological clinic in Finland since 1985, followed-up for mortality until 31 December 2008. The dataset contains the following variables:

## Loading the packages and the data

Load the R packages , and needed in this exercise.

Read the datafile {oralca2.txt} from a website, whose precise address will be given in the practical, into an R data frame named .

Look at the head, structure and the summary of the data frame. Using function count the numbers of censorings as well as deaths from oral cancer and other causes, respectively, from the variable.

orca <-   
 read.csv("C:/Users/janne.pitkaniemi/Projects/TRE2018/oralca2.txt",   
 sep="")  
head(orca)

sex age stage time event  
1 Male 65.42274 unkn 5.081 0  
2 Female 83.08783 III 0.419 1  
3 Male 52.59008 II 7.915 2  
4 Male 77.08630 I 2.480 2  
5 Male 80.33622 IV 2.500 1  
6 Female 82.58132 IV 0.167 2

## Estimate cumulative incidence using competing risks

1. Use KM-estimator for oral cancer and other deaths separately and estimate 5 year mortality

km1<- survfit( Surv( time, 1\*(event==1)) ~ 1,  
 data = orca)  
res<-summary(km1)  
index<-which(floor(res$time)==5)[1]-1  
  
cat("Probablity of dying to oral cancer before 5th year and 95%CI",  
 1-res$surv[index],"(",  
 1-res$lower[index],";",  
 1-res$upper[index],") \n" )

Probablity of dying to oral cancer before 5th year and 95%CI 0.3342919 ( 0.385453 ; 0.2788716 )

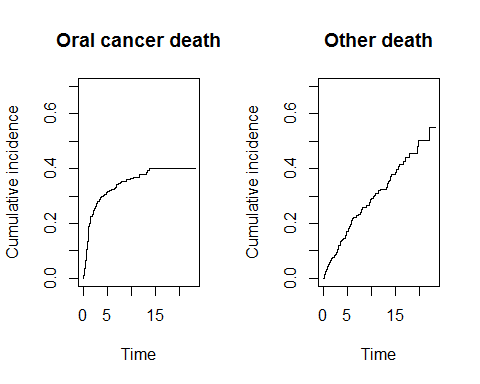
km1<- survfit( Surv( time, 1\*(event==2)) ~ 1,  
 data = orca)  
res<-summary(km1)  
index<-which(floor(res$time)==5)[1]-1  
  
cat("Probablity of dying to oral cancer before 5th year and 95%CI",  
 1-res$surv[index],"(",  
 1-res$lower[index],";",  
 1-res$upper[index],") \n" )

Probablity of dying to oral cancer before 5th year and 95%CI 0.2262094 ( 0.2784785 ; 0.1701539 )

## Estimate cumulative incidence using competing risks

1. Type help(plotCIF) and look at the help of plotting AJ-estimator for CIF. Then plot CIF for oral cancer data for deaths due to cancer and other causes.

par(mfrow=c(1,2))  
cif<- survfit( Surv( time, event, type="mstate") ~ 1,  
 data = orca)  
 plotCIF(cif, 1, main = "Oral cancer death",  
 col=1, ylim = c(0, 0.7) )  
 plotCIF(cif, 2, main= "Other death",  
 col=1, ylim = c(0, 0.7) )



Print summary of cif and find 5-year CIF for oral and other cause death

res<-summary(cif)  
index<-which(floor(res$time)==5)[1]-1  
probs<-res$pstate[index,]  
  
cat("Probablity of dying to oral cancer before 5th year and 95%CI",  
 probs[1],"(",  
 res$lower[index,1],";",  
 res$upper[index,1],") \n" )

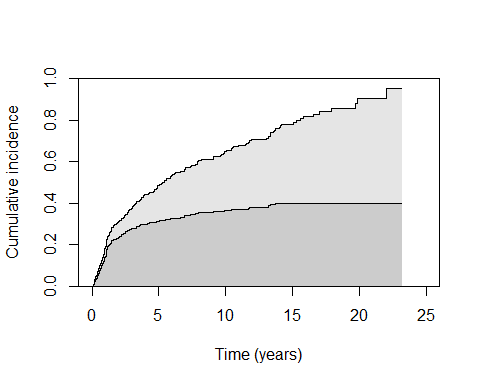
Probablity of dying to oral cancer before 5th year and 95%CI 0.3130833 ( 0.2611398 ; 0.3613751 )

cat("Probablity of dying to other causes before 5th year and 95%CI",  
 probs[2],"(",  
 res$lower[index,2],";",  
 res$upper[index,2],")" )

Probablity of dying to other causes before 5th year and 95%CI 0.1720909 ( 0.1294551 ; 0.2126386 )

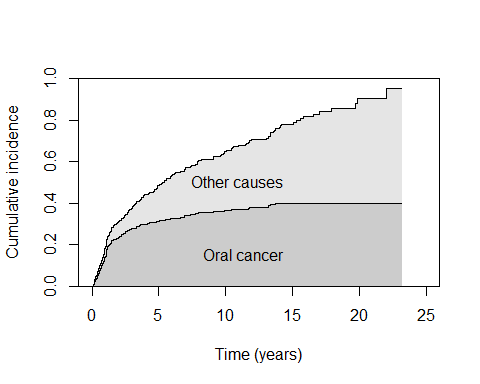
1. Type help(stackedCIF) for plotting stacked version of the CIF for both causes of death.

par(mfrow=c(1,1))  
stackedCIF(cif, colour = c("gray80", "gray90"),  
 main = "", xlab="Time (years)",xlim=c(0,25) )



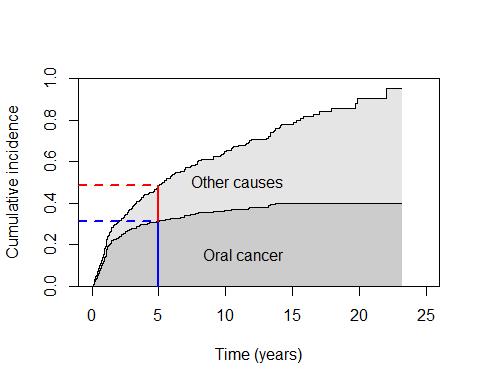
Add text indicating areas under the curve

par(mfrow=c(1,1))  
stackedCIF(cif, colour = c("gray80", "gray90"),  
 main = "", xlab="Time (years)",xlim=c(0,25) )  
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other causes", pos = 2)



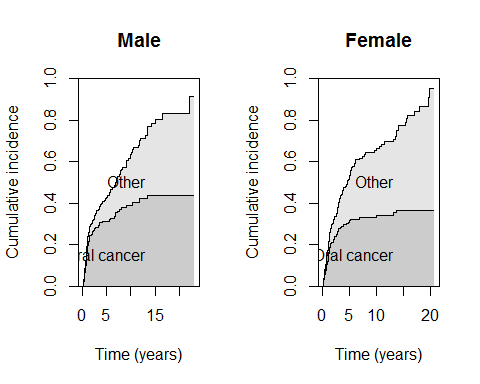
Add reference lines for 5-year mortalities.

par(mfrow=c(1,1))  
stackedCIF(cif, colour = c("gray80", "gray90"),  
 main = "", xlab="Time (years)",xlim=c(0,25) )  
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other causes", pos = 2)  
  
res<-summary(cif)  
index<-which(floor(res$time)==5)[1]-1  
probs<-res$pstate[index,]  
segments(5, 0, 5, probs[1], col= 'blue',lwd=2)  
segments(5, probs[1], 5, probs[1]+probs[2], col= 'red',lwd=2)  
  
segments(-1, probs[1], 5, probs[1], col= 'blue',lwd=2,lty=2)  
segments(-1, probs[1]+probs[2], 5, probs[1]+probs[2], col= 'red',lwd=2,lty=2)



1. Plot stacked CIF for males and females for oral and all cause mortality.

orca$nsex<-1\*(orca$sex=="Male")  
  
par(mfrow=c(1,2))  
cif<- survfit( Surv( time, event, type="mstate") ~ nsex,  
 data = orca)  
stackedCIF(cif, group=1, colour = c("gray80", "gray90"),  
 main = "Male", xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)  
  
stackedCIF(cif, group=2, colour = c("gray80", "gray90"),  
 main = "Female", xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)



1. Explore the effect of gender to death causes using cox model and report hazard ratio for males vs females.

cat("Oral cancer mortality")

Oral cancer mortality

ci.exp(coxph(Surv( time, event==1) ~ nsex, data = orca))

exp(Est.) 2.5% 97.5%  
nsex 0.896679 0.6283736 1.279546

cat("Other cause mortality")

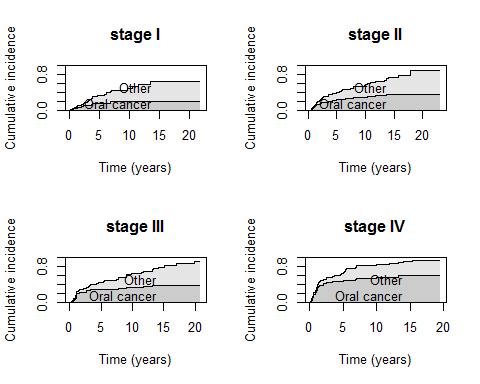
Other cause mortality

ci.exp(coxph(Surv( time, event==2) ~ nsex, data = orca))

exp(Est.) 2.5% 97.5%  
nsex 1.4963 1.010495 2.215658

* 1. Look at oral cancer and other cause mortality by stage

par(mfrow=c(2,2))  
cif<- survfit( Surv( time, event, type="mstate") ~ stage,  
 data = orca)  
  
stackedCIF(cif, group=1, colour = c("gray80", "gray90"),  
 main = "stage I", xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)  
  
stackedCIF(cif, group=2, colour = c("gray80", "gray90"),  
 main = "stage II", xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)  
  
stackedCIF(cif, group=3, colour = c("gray80", "gray90"),  
 main = "stage III", xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)  
  
stackedCIF(cif, group=4, colour = c("gray80", "gray90"),  
 main = "stage IV", xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)



cat("Oral cancer mortality")

Oral cancer mortality

ci.exp(coxph(Surv( time, event==1) ~ stage, data = orca))

exp(Est.) 2.5% 97.5%  
stageII 1.770772 0.8263026 3.794776  
stageIII 2.071560 0.9667494 4.438958  
stageIV 4.547819 2.1923553 9.433991  
stageunkn 2.764416 1.2938467 5.906415

cat("Other cause mortality")

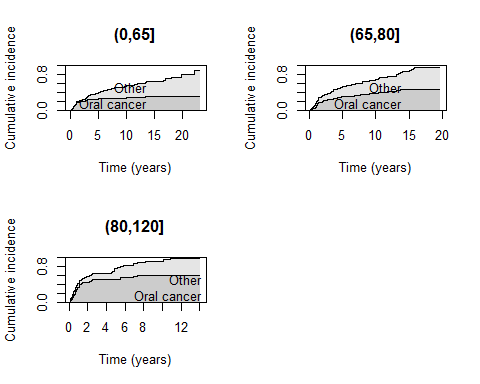
Other cause mortality

ci.exp(coxph(Surv( time, event==2) ~ stage, data = orca))

exp(Est.) 2.5% 97.5%  
stageII 0.9918711 0.5314690 1.851111  
stageIII 1.1944220 0.6378248 2.236733  
stageIV 1.7030515 0.8779929 3.303426  
stageunkn 1.4490857 0.7405324 2.835594

1. Look at oral cancer and other cause mortality by the following age groups: (0,65] , (65,80] , (80,120]

orca$ageg<-cut(orca$age,breaks = c(0,65,80,120))  
  
par(mfrow=c(2,2))  
cif<- survfit( Surv( time, event, type="mstate") ~ ageg,  
 data = orca)  
  
stackedCIF(cif, group=1, colour = c("gray80", "gray90"),  
 main = levels(orca$ageg)[1], xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)  
  
stackedCIF(cif, group=2, colour = c("gray80", "gray90"),  
 main = levels(orca$ageg)[2], xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)  
  
stackedCIF(cif, group=3, colour = c("gray80", "gray90"),  
 main = levels(orca$ageg)[3], xlab="Time (years)" )   
text( 15, 0.15, "Oral cancer", pos = 2)  
text( 15, 0.5, "Other", pos = 2)



## Time dependent effects- Heart transplant data

Look at the mortality related to the heart transplant using pre transplant time as refere and split the follow-up as in the lectures. Essentially we are redoing the analysis in the lecture for practice. In this case we have predefined time intervals of mortality risk.

First a naive Cox analysis with age, year and surgery as covarites and treatment as the factor of interest.

ci.lin(coxph(Surv(start, stop, event) ~  
 age + year + surgery + transplant,  
data=heart),Exp=TRUE)[,c(5:7)]

exp(Est.) 2.5% 97.5%  
age 1.0275390 1.0002875 1.0555330  
year 0.8638585 0.7524197 0.9918021  
surgery 0.5287657 0.2574423 1.0860419  
transplant1 0.9898016 0.5351550 1.8306980

Use the data called jasa as the heart data, because this version of the data consists also dates for accepting into the program (accept.dt), transplant date (tx.date), and follow-up time (futime).

heart.Lx <-  
Lexis(entry=list(age.time=age,   
 fu.time=0,  
tx.time=decimal\_date(tx.date)-decimal\_date(accept.dt)),  
duration=futime/365,   
exit.status=as.character(fustat),  
data=subset(jasa, is.na(tx.date) | fu.date != tx.date))

NOTE: entry.status has been set to "0" for all.

Warning in Lexis(entry = list(age.time = age, fu.time = 0, tx.time = decimal\_date(tx.date) - : Dropping 1 rows with duration of follow up < tol

head(heart.Lx)

age.time fu.time tx.time lex.dur lex.Cst lex.Xst lex.id  
1 30.84463 0 NA 0.134246575 0 1 1  
2 51.83573 0 NA 0.013698630 0 1 2  
3 54.29706 0 0.00000000 0.041095890 0 1 3  
4 40.26283 0 0.09562842 0.104109589 0 1 4  
5 20.78576 0 NA 0.046575342 0 1 5  
6 54.59548 0 NA 0.005479452 0 1 6  
 birth.dt accept.dt tx.date fu.date fustat surgery age  
1 1937-01-10 1967-11-15 <NA> 1968-01-03 1 0 30.84463  
2 1916-03-02 1968-01-02 <NA> 1968-01-07 1 0 51.83573  
3 1913-09-19 1968-01-06 1968-01-06 1968-01-21 1 0 54.29706  
4 1927-12-23 1968-03-28 1968-05-02 1968-05-05 1 0 40.26283  
5 1947-07-28 1968-05-10 <NA> 1968-05-27 1 0 20.78576  
6 1913-11-08 1968-06-13 <NA> 1968-06-15 1 0 54.59548  
 futime wait.time transplant mismatch hla.a2 mscore reject  
1 49 NA 0 NA NA NA NA  
2 5 NA 0 NA NA NA NA  
3 15 0 1 2 0 1.11 0  
4 38 35 1 3 0 1.66 0  
5 17 NA 0 NA NA NA NA  
6 2 NA 0 NA NA NA NA

Next, in order to make the cut points for subject separately we create three new variables for line

heart.Lx <- within(heart.Lx, {cut.0 <- tx.time;   
 cut.1 <- tx.time+0.2;  
 cut.2 <- tx.time+0.4} );

head(heart.Lx)

age.time fu.time tx.time lex.dur lex.Cst lex.Xst lex.id  
1 30.84463 0 NA 0.134246575 0 1 1  
2 51.83573 0 NA 0.013698630 0 1 2  
3 54.29706 0 0.00000000 0.041095890 0 1 3  
4 40.26283 0 0.09562842 0.104109589 0 1 4  
5 20.78576 0 NA 0.046575342 0 1 5  
6 54.59548 0 NA 0.005479452 0 1 6  
 birth.dt accept.dt tx.date fu.date fustat surgery age  
1 1937-01-10 1967-11-15 <NA> 1968-01-03 1 0 30.84463  
2 1916-03-02 1968-01-02 <NA> 1968-01-07 1 0 51.83573  
3 1913-09-19 1968-01-06 1968-01-06 1968-01-21 1 0 54.29706  
4 1927-12-23 1968-03-28 1968-05-02 1968-05-05 1 0 40.26283  
5 1947-07-28 1968-05-10 <NA> 1968-05-27 1 0 20.78576  
6 1913-11-08 1968-06-13 <NA> 1968-06-15 1 0 54.59548  
 futime wait.time transplant mismatch hla.a2 mscore reject cut.2  
1 49 NA 0 NA NA NA NA NA  
2 5 NA 0 NA NA NA NA NA  
3 15 0 1 2 0 1.11 0 0.4000000  
4 38 35 1 3 0 1.66 0 0.4956284  
5 17 NA 0 NA NA NA NA NA  
6 2 NA 0 NA NA NA NA NA  
 cut.1 cut.0  
1 NA NA  
2 NA NA  
3 0.2000000 0.00000000  
4 0.2956284 0.09562842  
5 NA NA  
6 NA NA

heart.Lx.cut <- mcutLexis(heart.Lx, timescale="fu.time",  
wh=c("cut.0","cut.1","cut.2"),  
new.states=c("a","b","c"))

Look at the variable lex.Cst whic indicates the state at the entry to follow-up period.

table(heart.Lx.cut$lex.Cst)

0 1 a a-b a-b-c   
 99 0 68 38 35

Do the cox regression using lex.Cst as factor covariate in the analysis.

print(format(as.data.frame(ci.exp(coxph(   
 Surv(fu.time, fu.time + lex.dur, lex.Xst==1)  
 ~ factor(lex.Cst),data=heart.Lx.cut))),digits=3))

exp(Est.) 2.5% 97.5%  
factor(lex.Cst)a 1.371 0.7355 2.56  
factor(lex.Cst)a-b 0.252 0.0501 1.26  
factor(lex.Cst)a-b-c 0.528 0.1817 1.53

If we want to adjust for the age at the biginning of the follow-up period we need the add age.time in the analysis.Let look at subject with lex.id =66.

heart.Lx.cut[heart.Lx.cut$lex.id==66,]

age.time fu.time tx.time lex.dur lex.Cst lex.Xst lex.id  
156 19.55099 0.0000000 0.1530055 0.1530055 0 a 66  
157 19.70400 0.1530055 0.3060109 0.2000000 a a-b 66  
158 19.90400 0.3530055 0.5060109 0.2000000 a-b a-b-c 66  
159 20.10400 0.5530055 0.7060109 0.2250767 a-b-c 1 66  
 birth.dt accept.dt tx.date fu.date fustat surgery age  
156 1952-09-03 1972-03-23 1972-05-18 1973-01-01 1 0 19.55099  
157 1952-09-03 1972-03-23 1972-05-18 1973-01-01 1 0 19.55099  
158 1952-09-03 1972-03-23 1972-05-18 1973-01-01 1 0 19.55099  
159 1952-09-03 1972-03-23 1972-05-18 1973-01-01 1 0 19.55099  
 futime wait.time transplant mismatch hla.a2 mscore reject cut.2  
156 284 56 1 3 0 1.02 0 0.5530055  
157 284 56 1 3 0 1.02 0 0.5530055  
158 284 56 1 3 0 1.02 0 0.5530055  
159 284 56 1 3 0 1.02 0 0.5530055  
 cut.1 cut.0  
156 0.3530055 0.1530055  
157 0.3530055 0.1530055  
158 0.3530055 0.1530055  
159 0.3530055 0.1530055

print(format(as.data.frame(ci.exp(coxph(   
 Surv(fu.time, fu.time + lex.dur, lex.Xst==1)  
 ~ factor(lex.Cst)+age.time,data=heart.Lx.cut))),digits=5))

exp(Est.) 2.5% 97.5%  
factor(lex.Cst)a 1.19981 0.636977 2.2600  
factor(lex.Cst)a-b 0.21164 0.042007 1.0663  
factor(lex.Cst)a-b-c 0.38181 0.124643 1.1696  
age.time 1.03453 1.004713 1.0652

Use some other variable that you find interesting and try to add it as a covariate of stratify the analysis.