finalsim.R

sabrina

2020-12-18

```
#########################################################################
# simulation if cox PH model is assumed, with some continuous covariates;
# baseline event time is assumed to follow Weibull/exponential distribution
# indepedent (uniform) censoring and Right censoring
#########################################################################
library(MASS)


sim_cox<- function(N,lambda0, beta, censor.right)
{
  # N = Total sample size
  # beta = PH coefficients
  # lambda0 = rate parameter of the exponential distribution for baseline


  #gender=0 is female and gender =1 is male
  gender <- sample(x=c(0, 1), size=N, replace=TRUE, prob=c(0.379, 0.621))
  #marriage=1: married, 2: never married, 3:Widowed, 4:divorced, 5:others
  marital<-sample(x=c( 1, 2, 3, 4, 5), size=N, replace=TRUE, prob=c(0.6, 0.166, 0.103, 0.058, 0.073))
  #race =1: white, 2: API, 3: black, 4: others
  race<-sample(x=c( 1, 2, 3, 4), size=N, replace=TRUE, prob=c(0.884, 0.072, 0.025, 0.02))
  #tumorLoc= 1: Mobile spine, 2:sacrum
  tumorLoc<-sample(x=c( 1, 2), size=N, replace=TRUE, prob=c(0.443, 0.557))
  # #patho =1: classic, 2: chondroid, 3: dedifferentiated
  patho<-sample(x=c( 1, 2, 3), size=N, replace=TRUE, prob=c(0.976, 0.016, 0.008))
  # #diagtime =1 1974-1983 =2: 1984-1993 =3: 1994-2003 =4: 2004-2013
  diagtime<-sample(x=c( 1, 2, 3, 4), size=N, replace=TRUE, prob=c(0.089, 0.134, 0.262, 0.515))
  # #therapy =1: radiology =2:Surgery alone =3: Surgery and radiotherapy alone =4: neither surgery nor
  therapy<-sample(x=c( 1, 2, 3, 4), size=N, replace=TRUE, prob=c(0.1, 0.1, 0.7, 0.1))

  # generate continuous covariates, mutually indepedent
  #going to assume zero correlation between the two

  X = mvrnorm(N,mu=c(60.3,82.5),Sigma=matrix(c(17.1,0,0,74.2),2,2))
  age=X[,1]
  tumorSize=X[,2]


  # initial data set
  initial<-data.frame(id=1:N,
                      Gender=gender,
                      Marital = marital,
```

```
                    Race= race,
                    TumorLocation= tumorLoc,
                    Pathology=patho,
                    Diagtime= diagtime,
                    Therapy= therapy,
                    Age=age,
                    TumorSize=tumorSize)


#sacrum location =1 if tumorloc = 2
sacrum<- c(1:N)*0;
for (i in 1:N) {
  if(initial$TumorLocation[i]==2)
  {
    sacrum[i]<- 1;
  }
}

#surgery  =1 if therapy = 2
surgery<- c(1:N)*0;
for (i in 1:N) {
  if(initial$Therapy[i]==2)
  {
    surgery[i]<- 1;
  }
}

#radiotherapy  =1 if therapy = 1
rad<- c(1:N)*0;
for (i in 1:N) {
  if(initial$Therapy[i]==1)
  {
    rad[i]<- 1;
  }
}

#RS  =1 if therapy = 3
RS<- c(1:N)*0;
for (i in 1:N) {
  if(initial$Therapy[i]==3)
  {
    RS[i]<- 1;
  }
}

#classic  =1 if pathology = 1
classic<- c(1:N)*0;
for (i in 1:N) {
  if(initial$Pathology[i]==1)
  {
    classic[i]<- 1;
  }
}
```

```r
  #chondroid  =1 if pathology=2
  chon<- c(1:N)*0;
  for (i in 1:N) {
    if(initial$Pathology[i]==2)
    {
      chon[i]<- 1;
    }
  }

  initial<-cbind(initial, sacrum, surgery, rad, RS, classic, chon)

  # generate underlying event time
  # T <- rweibull(n=N, shape=1, scale = lambda0*exp(beta[1]*young+beta[2]*surgery+beta[3]*sacrum))
  #

  T <- rweibull(n=N, shape=1, scale = lambda0*exp(beta[1]*age+beta[2]*sacrum
                                                  +beta[3]*surgery + beta[4]*RS + beta[5]*classic))

  #mean(X)
  #rexp(n=N, rate=lambda0*exp(beta*A))

  # censoring times
  ctime = runif(N, min=0, max=censor.right)

  # follow-up times and event indicators
  # time= c(1:N)*0
  # for(i in 1:N)
  # { if(initial$Diagtime[i]==4)
  # {
  #   time[i]<- pmin(T, ctime, 160)
  # }
  #   else
  #   {
  #     time[i] <- pmin(T, ctime, censor.right)
  #   }
  #
  # }

  time<- pmin(T, ctime, censor.right)

  censor <- as.numeric(T>ctime | T>censor.right)
  finalData<-cbind(initial, time, censor)

  return(finalData)

}
#median follow up time was 52 months so for lambda0=200 we get median time approx 52.
#latest censor time was 480 months so censor.right=480

finalSimP2<-sim_cox(N=765, lambda0=200, beta=c(-log(1.052), -log(0.668),-log(0.288), -log(0.524),-log(0

#data check
median(finalSimP2$time)
```

```
## [1] 53.11406
```

```
mean(finalSimP2$Age)
```

```
## [1] 60.36272
```

```
mean(finalSimP2$TumorSize)
```

```
## [1] 82.62102
```

```
sum(finalSimP2$Gender==0)
```

```
## [1] 287
```

```
sum(finalSimP2$time< 60)
```

```
## [1] 412
```

```
sum(finalSimP2$time< 120)
```

```
## [1] 592
```

```
sum(finalSimP2$Diagtime==4)
```

```
## [1] 389
```

```
write.csv(finalSimP2,
          file="/Users/sabrina/Desktop/Fall 2020/BST222/Project 2/finalsimp2.csv", row.names = FALSE)
```