

# Analysis of Time-to-Event Data: Study Sheet 3

Submission: René-Marcel Kruse, Doctorial Candidate

## Exercise 1:

In **R** make use of the function `rgamma()` to generate 250 event times from a Gamma distribution with shape parameter 1.5 and scale parameter 0.75. Then, generate censored observations according to the following censoring schemes:

(a)

For every individual  $i = 1, \dots, n$  there is a fixed observation period  $c_i$  (type I censoring). Assume that the fixed observation period is the same for all individuals and set  $c = 1, 3, 5$ .

Answer:

Censor times

```
t1 <- pmin(true.t,1)
t2 <- pmin(true.t,3)
t3 <- pmin(true.t,5)
```

deltas

```
d1 <- 1*(true.t<=1)
d2 <- 1*(true.t<=3)
d3 <- 1*(true.t<=5)
```

Survival

```
surv1 <- Surv(t1,d1)
surv2 <- Surv(t2,d2)
surv3 <- Surv(t3,d3)
```

KM-estimator

```
km1 <- survfit(surv1 ~ 1,type="kaplan-meier")
km2 <- survfit(surv2 ~ 1,type="kaplan-meier")
km3 <- survfit(surv3 ~ 1,type="kaplan-meier")
```

Nelson-Ahlen Estimator

```
na1 <- NA_estimator(t1,d1)
na2 <- NA_estimator(t2,d2)
na3 <- NA_estimator(t3,d3)
```

Sum of total events

```
sum(d1)
```

```
## [1] 88
```

```
sum(d2)
```

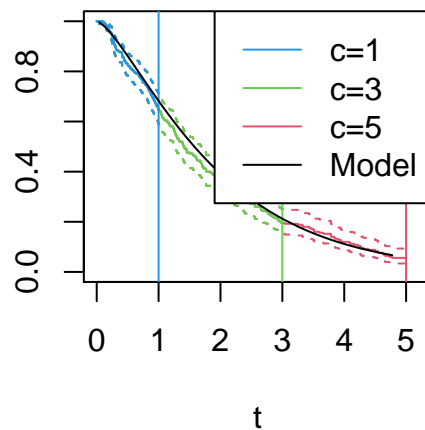
```
## [1] 201
```

```
sum(d3)
```

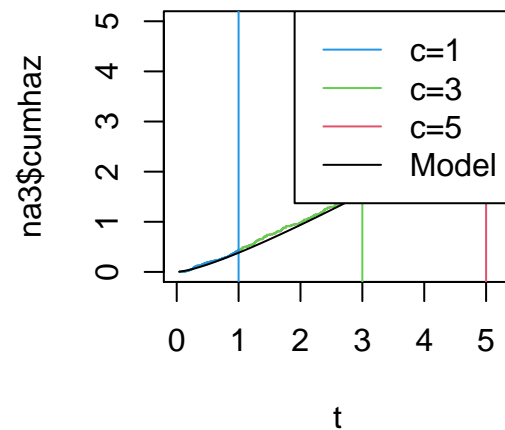
```
## [1] 236
```

Answer:

### Kaplan–Meier estimator



### Nelson–Aalen estimator



(b)

The study continues until the failure of the first  $k$  individuals, where  $k$  is some predetermined integer ( $k < n$ ) (type II censoring). Set  $k = 100, 150, 200$ .

### Censoring Times

```
t1 <- c(true.t[1:100], rep(true.t[100], n-100))
t2 <- c(true.t[1:150], rep(true.t[150], n-150))
t3 <- c(true.t[1:200], rep(true.t[200], n-200))
```

### Deltas

```
d1 <- c(rep(1,100), rep(0,n-100))
d2 <- c(rep(1,150), rep(0,n-150))
d3 <- c(rep(1,200), rep(0,n-200))
```

## Survivals

```
surv1 <- Surv(t1,d1)
surv2 <- Surv(t2,d2)
surv3 <- Surv(t3,d3)
```

## Kaplan-Meier Estimator

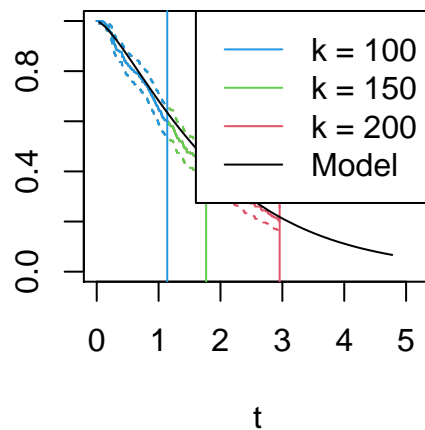
```
km1 <- survfit(surv1~1,type="kaplan-meier")
km2 <- survfit(surv2~1,type="kaplan-meier")
km3 <- survfit(surv3~1,type="kaplan-meier")
```

## Nelson-Aalen Estimator

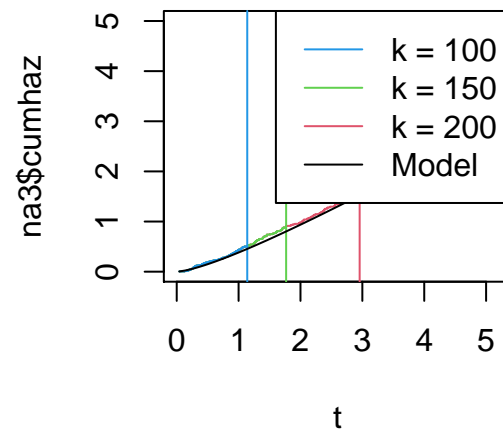
```
na1 <- NA_estimator(t1,d1)
na2 <- NA_estimator(t2,d2)
na3 <- NA_estimator(t3,d3)
```

Answer:

### Kaplan-Meier estimator



### Nelson-Aalen estimator



(c)

Each unit has associated with it a potential censoring time  $C_i$  and a potential lifetime  $T_i$ , which are assumed to be independent random variables (type III censoring or random censoring). Assume that the censoring times are exponentially distributed with parameter  $\lambda = 0.25, 0.5, 1$ .

## Censoring times

```
c1 <- rexp(n,rate=0.25)
t1 <- pmin(true.t,c1)
c2 <- rexp(n,rate=0.5)
```

```
t2    <- pmin(true.t,c1)
c3    <- rexp(n,rate=1)
t3    <- pmin(true.t,c1)
```

## Deltas

```
d1 <- 1*(true.t<=c1)
d2 <- 1*(true.t<=c2)
d3 <- 1*(true.t<=c3)
```

## Survivals

```
surv1 <- Surv(t1,d1)
surv2 <- Surv(t2,d2)
surv3 <- Surv(t3,d3)
```

## Sum of total events

```
sum(d1)
```

```
## [1] 159
```

```
sum(d2)
```

```
## [1] 126
```

```
sum(d3)
```

```
## [1] 78
```

## KM estimator

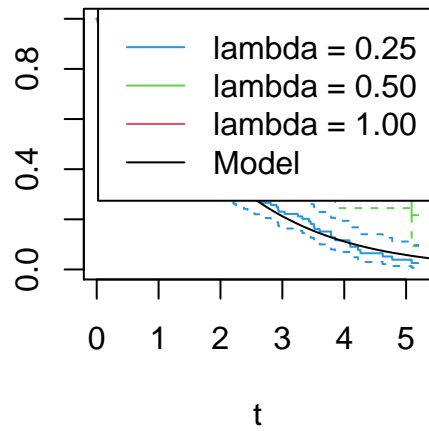
```
km1 <- survfit(surv1~1,type="kaplan-meier")
km2 <- survfit(surv2~1,type="kaplan-meier")
km3 <- survfit(surv3~1,type="kaplan-meier")
```

## NA estimator

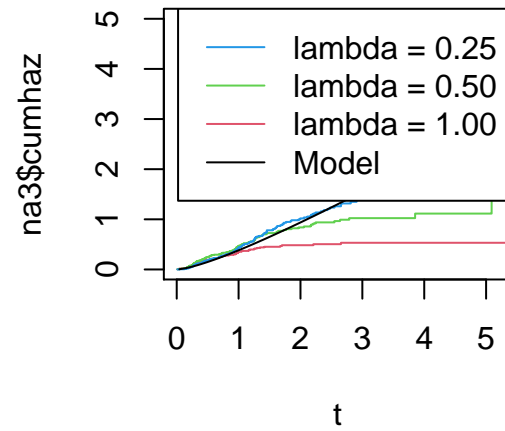
```
na1 <- NA_estimator(t1,d1)
na2 <- NA_estimator(t2,d2)
na3 <- NA_estimator(t3,d3)
```

Answer:

**Kaplan–Meier estimator**



**Nelson–Aalen estimator**



## Exercise 2:

The file `melanoma.dat`, which is stored along with a description in the Stud.IP folder “Data”, contains data of a clinical study at the Department of Plastic Surgery, University Hospital of Odense, Denmark. 205 patients were observed during the years 1962 to 1977 after a skin cancer operation (removal of the tumor and the surrounding skin), either up to the end of the year 1977 or their early death.

(a)

Write a function `lifetable <- function(time,delta,grid){...}` in **R** to implement the life-table method. The argument `time` corresponds to the event times of the  $n$  individuals, `delta` is a censoring indicator and `grid` contains the points used for discretization of the time axis. For a given set of data, the function should return the conditional probability of experiencing the event in each interval, the estimated survival function at the start of each interval, and the estimated density and hazard rate at the midpoint of each interval.

Answer: ... Fortunately still had the solutions from the year before last

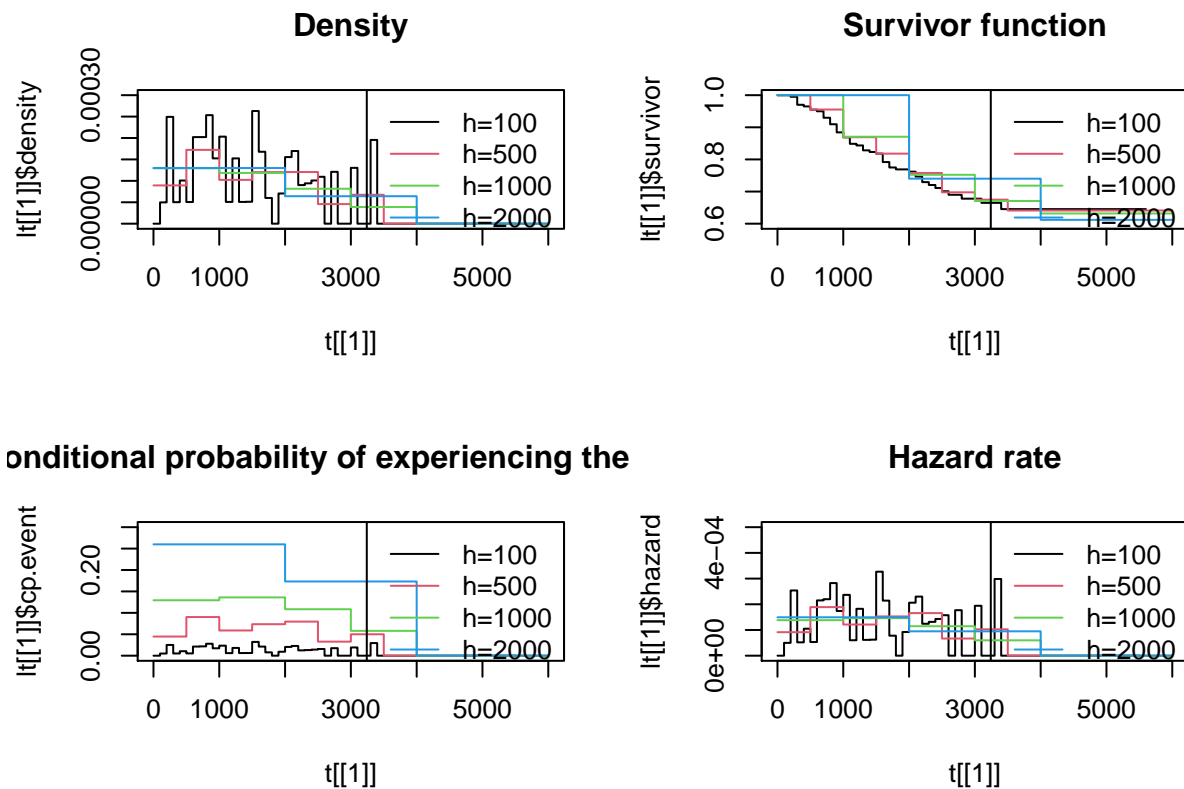
```
lifetable <- function(time,delta,grid)
{
  q <- length(grid)-1
  n <- length(time)
  d <- numeric(q)
  Y <- rep(n,q+1)
  w <- numeric(q)
  h <- numeric(q)

  for(k in 1:q)
  {
    d[k] <- sum(delta[time>=grid[k] & time<grid[k+1]])
    w[k] <- length(delta[time>=grid[k] & time<grid[k+1]])-d[k]
    Y[k+1] <- Y[k] - d[k] - w[k]
    h[k] <- grid[k+1] - grid[k]
  }
  cp.event <- survivor <- density <- hazard <- numeric(q)
  cp.event <- d / (Y[-(q+1)] - w/2)
  survivor <- cumprod(1-cp.event)
  density <- ( c(1,survivor[-q]) - survivor ) / h
  hazard <- 2*cp.event / (h * (2-cp.event))
  cp.event <- c(cp.event,cp.event[q])
  survivor <- c(1,survivor)
  density <- c(density,density[q])
  hazard <- c(hazard,hazard[q])
  erg <- list(cp.event=cp.event,
             survivor=survivor,
             density=density,
             hazard=hazard)
  return(erg)
}
```

(b)

Apply your function written in (a) to the skin cancer data. Define an appropriate censoring indicator `delta` and use equidistant time intervals (`grid`) between 0 and 6000 with interval lengths 100, 500, 1000 and 2000. Display each quantity that is asked for in (a) graphically and interpret the results.

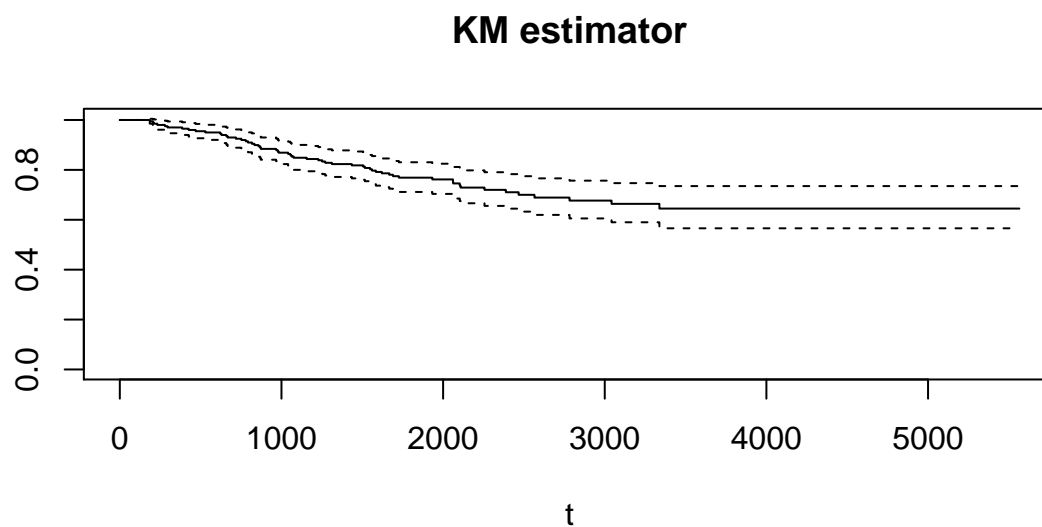
Answer:



(c)

In **R**, compute and plot the Kaplan-Meier estimate of the survivor function for the data of survival after malignant melanoma.

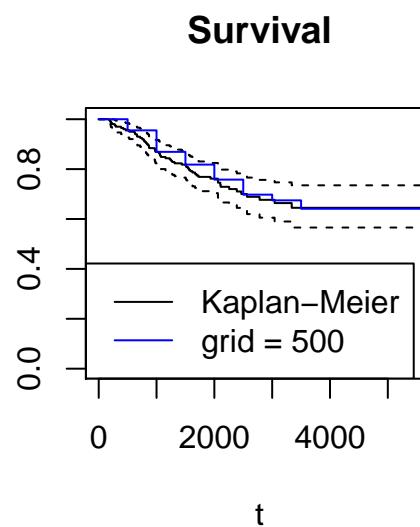
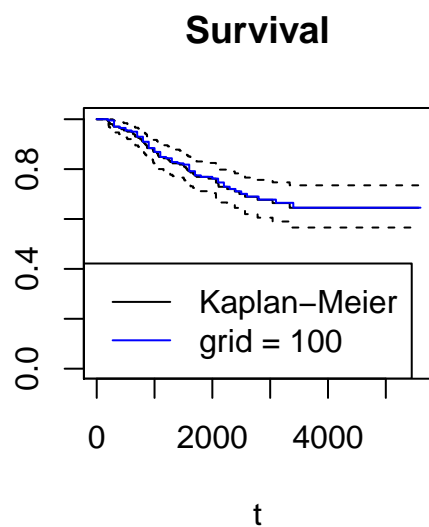
Answer:



(d)

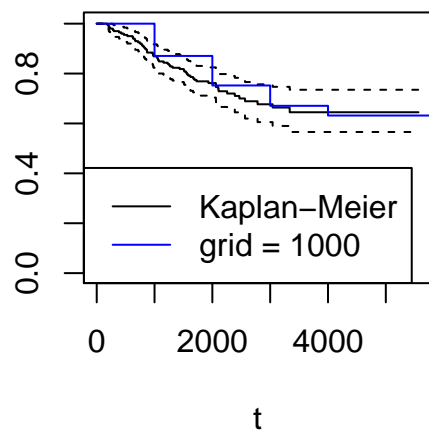
*Graphically compare the Kaplan-Meier estimate obtained in (c) with the estimate of the survivor function obtained by the life-table method.*

Answer:

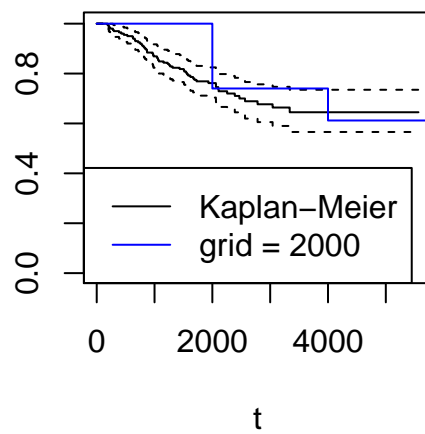




**Survival**



**Survival**



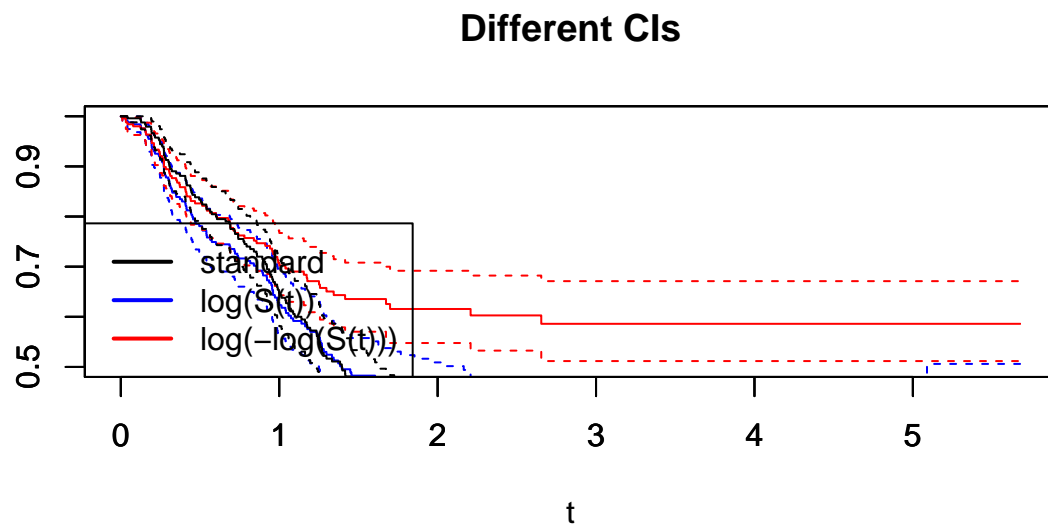
### Exercise 3:

Several proposals for computing pointwise confidence intervals for values of the survival function do exist.

(a)

Recall the data file `melanoma.dat` that was analysed in exercise 2. In **R**, use the `survfit()` function from the package `survival` to compute the Kaplan-Meier estimate of the survivor function and compare various pointwise 95% confidence intervals that are available in the option `conf.type`.

**Answer:**



(b)

Use the delta method that was introduced in the lectures to derive the asymptotic variance of  $\ln(-\ln(\hat{S}(t)))$ .

**Answer:**

The delta method is given as:

$$\text{Var}(f(x)) = \text{Var}(x) \cdot \left( \frac{\delta(f(x))}{\delta x} \right)^2$$

Applied to our problem we get:

$$\text{Var}\left(\ln(-\ln(\hat{S}(t)))\right) = \text{Var}\left(\ln(\hat{S}(t))\right) \cdot \left(\frac{\delta(\ln(-\ln(\hat{S}(t))))}{\delta(\ln(\hat{S}(t)))}\right)^2 \quad (1)$$

$$= \text{Var}\left(\ln(\hat{S}(t))\right) \cdot \frac{1}{\ln(\hat{S}(t))^2} \quad (2)$$

$$= \sum_{t_i \leq t} \frac{d_k}{n_k(n_k - d_k)} \cdot \frac{1}{\ln(\hat{S}(t))^2} \quad (3)$$

(c)

Based on the result obtained in (b) give a corresponding asymptotic  $100(1 - \alpha)\%$  confidence interval for  $S(t)$ .

**Answer:**

CI for  $\hat{S}(t)$  given by:

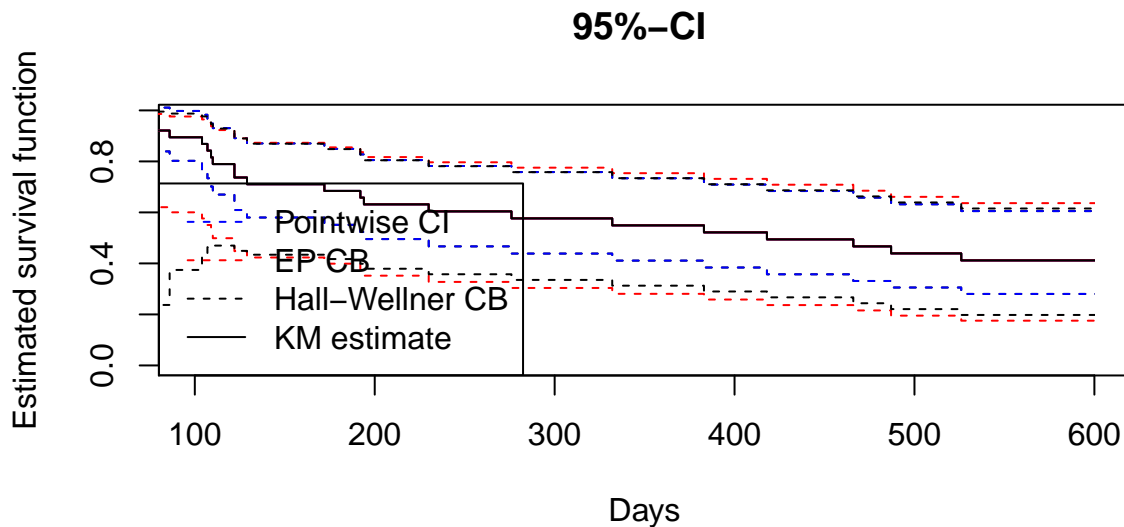
$$CI_{\pm} = \ln(-\ln(\hat{S}(t))) \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\text{Var}\left(\ln(-\ln(\hat{S}(t)))\right)}$$

## Exercise 4:

In this exercise we consider data on the efficiency of bone marrow transplants for patients with acute myelotic leukemia (AML) and acute lymphoblastic leukemia (ALL). The data are reported in section 1.3 and appendix D of Klein and Moeschberger (2003) and details of this study are given in Copelan et al. (1991). The data are contained in the data frame `bmt` in the **R** package `KMsurv`. Make use of `help(bmt)` to become acquainted with the data set.

(a)

We shall focus on the disease-free survival probabilities for the 38 patients with ALL. An individual is said to be disease free at a given time after transplant if that individual is alive without the recurrence of leukemia. In **R**, construct and graphically compare 95% pointwise confidence intervals, Equal-Precision (EP) confidence bands and Hall-Wellner confidence bands for the disease-free survival function in the range  $100 \leq t \leq 600$ . Hint: In the lecture, we used the function `confBands()` from the **R** package `OISurv` to construct confidence bands. Confidence bands can also be computed using the `km.ci()` function in the **R** package `km.ci`.



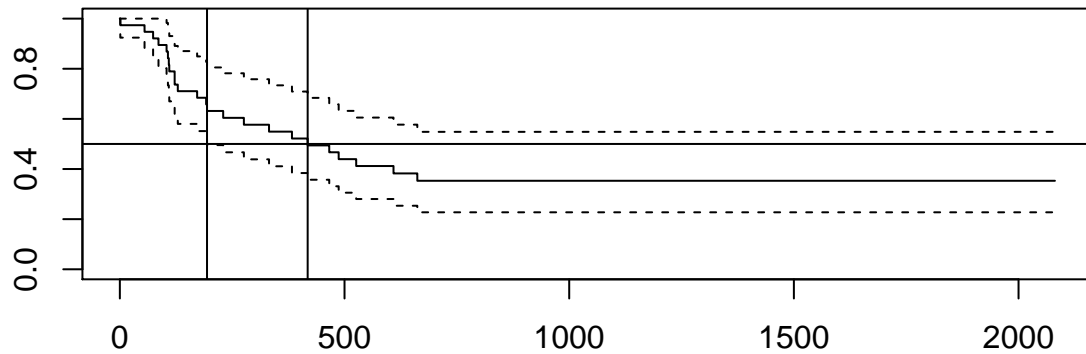
(b)

Estimate the mean disease-free survival time for the three disease groups.

```
## Call: survfit(formula = bmtSurv ~ 1)
##
##           n      events      *rmean *se(rmean)    median  0.95LCL  0.95UCL
##          38         24        899      146        418      194      NA
##    * restricted mean with upper limit = 2081

## Call: survfit(formula = bmtsurv2 ~ 1)
##
##           n      events      *rmean *se(rmean)    median  0.95LCL  0.95UCL
##          54         25       1549      151       2204      704      NA
##    * restricted mean with upper limit = 2569
```

```
## Call: survfit(formula = bmtsurv3 ~ 1)
##
##          n      events    *rmean *se(rmean)   median   0.95LCL   0.95UCL
##         45         34      792      158      183      115      456
##    * restricted mean with upper limit = 2640
```



(c)

*Estimate the median disease-free survival time for the three disease groups and construct a 95% confidence interval for the median.*

```
## 50
## 418

## 2.5 50.0 97.5
## 1 418 NA

## 2.5 50.0 97.5
## 35 2204 NA

## 2.5 50.0 97.5
## 16 183 NA
```