Survival Analysis
Lecture 4

Marta Fiocco[1,2] & Hein Putter [1]

(1) Department of Medical Statistics and Bioinformatics
Leiden University Medical Center
(2) Mathematical Institute Leiden University

LU
MC

Outline

KMsurv packages

CI S(t)

R code

Confidence intervals

Confidence bands

Mean and median $S(t)$
Example

Follow-up
Descriptives
Reverse Kaplan-Meier

KMsurv packages   **CI S(t)**   **R code**   **Confidence intervals**   **Confidence bands**   **Mean and median** $S(t)$ **Follow-up**

○○○○○○○○○   ○○○○○   ○○○   ○○○○○○○○   ○○○   ○○○○○○○○   ○○○○○○○○○○

Outline

KMsurv packages

CI S(t)

R code

Confidence intervals

Confidence bands

Mean and median $S(t)$
   Example

Follow-up
   Descriptives
   Reverse Kaplan-Meier

LU
MC

Outline

LU
MC

- ► A lot of functions (and data sets) for survival analysis are in the package survival
- ► we need to load it at first

```
> library(survival) #load library
> library(help=survival) # list of available functions and data sets
```

- ► data sets from Klein and Moeschberger

```
> library(KMsurv)
> library(help=KMsurv)
> # look at a specific data set in the library
> data(lung)
> # print the first 6 rows of data set lung
> head(lung)
  time death time2 death2
1  139     1   139      1
2  304     1   304      1
3  193     1   193      1
4  248     1   248      1
5   27     1    27      1
6  210     1   210      1
```

Some R code to begin with

- ▶ look at the time to event in the data set **aml** with function `Surv`

```
> ?Surv # see the help
> Surv(aml$time, aml$status)
 [1]   9   13  13+  18   23  28+  31   34  45+  48 161+   5    5
[16]  12  16+  23   27   30   33   43   45
```

- ▶ Kaplan-Meier and Nelson-Aalen estimator obtained with the R function `survfit()`
- ▶ estimate the distribution of lifetimes non-parametrically, based on right censored observations
- ▶ use Kaplan-Meier estimator; the R function to perform this is `survfit()`

```
> #estimate survival curve for aml data
> fit <- survfit(Surv(time, status) ~ x, data = aml)

> summary(fit)
Call: survfit(formula = Surv(time, status) ~ x, data = aml)

                x=Maintained
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    9     11       1    0.909  0.0867       0.7541        1.000
   13     10       1    0.818  0.1163       0.6192        1.000
   18      8       1    0.716  0.1397       0.4884        1.000
   23      7       1    0.614  0.1526       0.3769        0.999
   31      5       1    0.491  0.1642       0.2549        0.946
   34      4       1    0.368  0.1627       0.1549        0.875
   48      2       1    0.184  0.1535       0.0359        0.944
```

▶ output from `summary(fit)`

```
             x=Nonmaintained
time n.risk n.event survival std.err lower 95% CI upper 95% CI
   5     12       2   0.8333  0.1076       0.6470        1.000
   8     10       2   0.6667  0.1361       0.4468        0.995
  12      8       1   0.5833  0.1423       0.3616        0.941
  23      6       1   0.4861  0.1481       0.2675        0.883
  27      5       1   0.3889  0.1470       0.1854        0.816
  30      4       1   0.2917  0.1387       0.1148        0.741
  33      3       1   0.1944  0.1219       0.0569        0.664
  43      2       1   0.0972  0.0919       0.0153        0.620
  45      1       1   0.0000     NaN           NA           NA
```
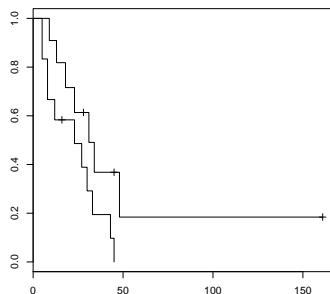
```
> # to know what has been estimated with survfit
> names(summary(fit))
 [1] "surv"          "time"         "n.risk"        "n.event"
 [5] "conf.int"      "type"         "table"         "std.err"
 [9] "lower"         "upper"        "strata"        "rmean.endtime"
[13] "call"
> # plot Kaplan-Meier estimator, ticks for the censored observations.
> plot(fit)
```

Standard errors in R

▶ look at the standard error from `survfit` and
  `summary.survit(fit)`

```
> fit$std.err
  [1] 0.09534626 0.14213381 0.19508758 0.24873417 0.24873417 0.33446777
  [7] 0.44181673 0.44181673 0.83378775 0.83378775 0.12909944 0.20412415
 [13] 0.24397502 0.24397502 0.30472470 0.37796447 0.47559487 0.62678317
 [19] 0.94491118           Inf
>
> sfit$std.err
  [1] 0.08667842 0.11629130 0.13966497 0.15263233 0.16419327 0.16266889
  [7] 0.15349275 0.10758287 0.13608276 0.14231876 0.14813006 0.14698618
 [13] 0.13871517 0.12187451 0.09186636           NaN
```

▶ the former is the standard error for the estimated
  cumulative hazards $\widehat{H}(t)$
▶ the latter is the standard error for the survival $\widehat{S}(t)$

Cumulative hazard function

► The cumulative hazard function and the survival function are related in the following way for continuous data:

$$S(t) = exp(-H(t))$$

► $H(t)$ may be obtained by the inverse transformation of the Kaplan-Meier estimate

$$\hat{H}(t) = -\log \hat{S}(t)$$

► Another method to estimate $H(t)$ is the Nelson-Aalen estimator:

$$\tilde{H}(t) = \sum_{t_i \leq t} \frac{d_i}{Y_i}, \quad \sigma_H^2 = \sum_{t_i \leq t} \frac{d_i}{Y_i^2}$$

► no function in the survival package calculates either form automatically, the object returned by `summary(survfit())` can be used to calculate the estimates

```
data(tongue); attach(tongue)
my.surv <- Surv(time[type==1], delta[type==1])
my.fit <- summary(survfit(my.surv ~ 1))
H.hat <- -log(my.fit$surv)
H.hat <- c(H.hat, tail(H.hat, 1))
```

► construct Nelson-Aalen

```
h.sort.of <- my.fit$n.event / my.fit$n.risk
H.tilde <- cumsum(h.sort.of)
H.tilde <- c(H.tilde, tail(H.tilde, 1))
plot(c(my.fit$time, 250), H.hat, xlab="time", ylab="cumulative hazard"
main="comparing cumulative hazards", ylim=range(c(H.hat, H.tilde)),
type="s")
points(c(my.fit$time, 250), H.tilde, lty=2, type="s")
legend("topleft", legend=c("H.hat","H.tilde"), lty=1:2, bty="n")
detach(tongue)
```

comparing cumulative hazards
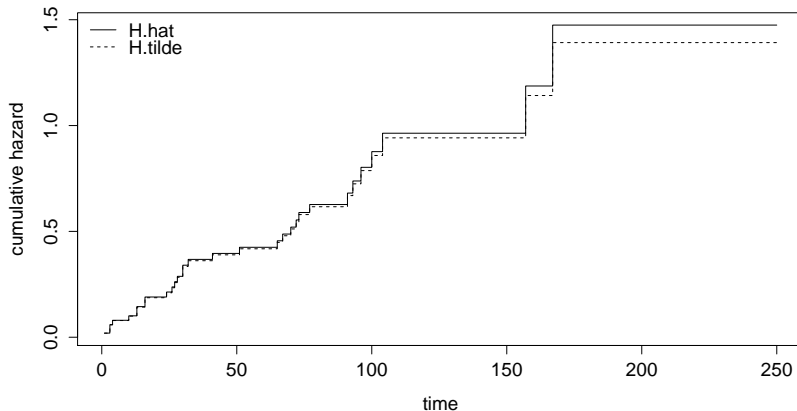
Outline

KMsurv packages

CI S(t)

R code

Confidence intervals

Confidence bands

Mean and median $S(t)$
Example

Follow-up
Descriptives
Reverse Kaplan-Meier

LU
MC

Pointwise confidence intervals for $\widehat{S}(t)$

- ▶ Product-Limit estimator provides an estimate of the survival
- ▶ the corresponding standard error provides some limited information about the precision of the estimate
- ▶ use these estimators to provide confidence intervals for the survival function at a fixed time t_0
- ▶ construct the intervals to assure, with a given confidence level $1 - \alpha$ that the true value of the survival function, at a predetermined time t_0 falls in this interval
- ▶ most common used confidence interval

$$\hat{S}(t_o) - Z_{1-\alpha/2}\sigma_S(t_o)\hat{S}(t_o), \, \hat{S}(t_o) + Z_{1-\alpha/2}\sigma_S(t_o)\hat{S}(t_o)$$

Pointwise confidence intervals for $\widehat{S}(t)$

- $\sigma_S^2(t) = \widehat{V}[\widehat{S}(t)]/\widehat{S}^2(t)$
- recall $\widehat{V}[\widehat{S}(t)] = \widehat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{Y_i(Y_i - d_i)}$
- $Z_{1-\alpha/2}$: $1 - \alpha/2$ percentile of a $N(0, 1)$ distribution
- this is the confidence interval constructed by most statistical packages
- better confidence intervals can be constructed by first transforming $\widehat{S}(t_0)$
- The "log"-transformed confidence interval is based on first finding a confidence interval for the log of the cumulative hazard function (called log-log transformed interval since the cumulative hazard function is the negative log of the survival function $H(t) = -\log(S(t))$)

Pointwise confidence intervals for $\widehat{S}(t)$

▶ the $100 \times (1 - \alpha)\%$ log-transformed confidence interval for the survival function at t_0 is given by

$$[\widehat{S}(t_0)^{1/\theta}, \widehat{S}(t_0)^{\theta}], \ \ \theta = \exp\{\frac{Z_{1-\alpha/2}\sigma_S^2(t_0)}{\log[\widehat{S}(t_0)]}\}$$

▶ the second transformation is an arcsine-square root transformation of the survival function; this yields the following $100 \times (1 - \alpha)\%$ confidence interval for the survival function

Second transformation for the confidence intervals

$$\sin^2\left\{\max\left[0, \arcsin(\hat{S}(t_o)^{1/2}) - 0.5 Z_{1-\alpha/2}\sigma_S(t_o)\left(\frac{\hat{S}(t_o)}{1-\hat{S}(t_o)}\right)^{1/2}\right]\right\}$$

$$\leq S(t_o) \leq \qquad\qquad (4.3.3)$$

$$\sin^2\left\{\min\left[\frac{\pi}{2}, \arcsin(\hat{S}(t_o)^{1/2}) + 0.5 Z_{1-\alpha/2}\sigma_S(t_o)\left(\frac{\hat{S}(t_o)}{1-\hat{S}(t_o)}\right)^{1/2}\right]\right\}.$$

Marta Fiocco [1,2] & Hein Putter[1]

"Plain" intervals

- confidence intervals can be computed on the scale of $S(t)$ as

$$\widehat{S}(t) \pm 1.96\sqrt{\text{var}[\widehat{S}(t)]}$$

- where the variance is computed with the Greenwood's formula
- it has been shown the inferiority of the "plain" intervals in the sense that intervals may extend below zero or above 1
- the plain scale interval has poor coverage properties
- therefore it is better to avoid them

Outline

KMsurv packages

CI S(t)

R code

Confidence intervals

Confidence bands

Mean and median $S(t)$
Example

Follow-up
Descriptives
Reverse Kaplan-Meier

Some R code

```
> library(survival)
> data(lung)
kfit <- survfit(Surv(time2,death2) ~1, data=lung)
> print(kfit)
Call: survfit(formula = Surv(time2, death2) ~ 1, data = lung)

records  n.max n.start  events  median 0.95LCL 0.95UCL
    25     25     25      13     210     144      NA
```

- ▶ the left-hand side of the formula declares the response variable to be a survival object
- ▶ the right-hand side ~ 1 is a null model,

```
> plot(kfit, col="blue") #ticks for the censored observations
# make the plot a bit nicer...
plot(kfit, mark.time=F, xscale=365.25, xlab="Years", ylab="Survival")
# in the plot command mark.time=F prevents "+" marks from being added
#at the censoring times
# if you want to see censored observations
plot(kfit, mark.time=T, xscale=365.25, xlab="Years", ylab="Survival")
```

LU
MC

R code

- ▸ estimate the survival as suggested by Breslow (1972)
  $\widetilde{S}(t) = \exp(-\widetilde{H}(t)$
  ```
  > ffit <- survfit(Surv(time2,death2) ~1, data=lung,
  type="fleming-harrington")
  ```

- ▸ plot the two estimates along with a 95% confidence interval for the K-M
  ```
  > plot(kfit, mark.time=F, xscale=365.25, xlab="Years",
  ylab="Survival")
  > lines(ffit,mark.time=F, xscale=365.25, col= "red")
  ```

LU
MC

Kaplan-Meier (dark line) and Nelson-Aalen estimates for the lung data set
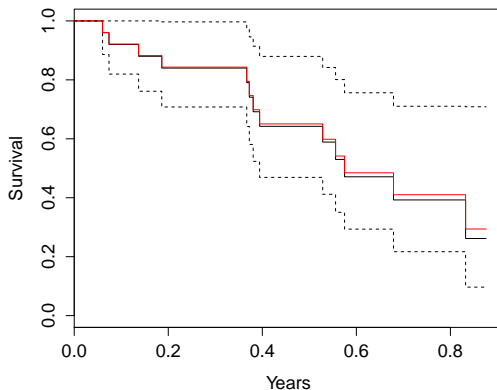
Outline

KMsurv packages

CI S(t)

R code

Confidence intervals

Confidence bands

Mean and median $S(t)$
    Example

Follow-up
    Descriptives
    Reverse Kaplan-Meier

LU
MC

Computation of confidence intervals

- use the estimated disease-free survival function and cumulative hazard rate for ALL patients in Section 1.3
  $\widehat{S}(365) = 0.5492$ $\widehat{V}[\widehat{S}(365)] = 0.0812^2$

$$\sigma_S^2(365) = (\frac{0.0812}{0.5492})^2 = 0.1479^2$$

- 95% confidence interval for the survival function at one year $0.5492 \pm 1.96 \times 0.1479 \times 0.5492 = (0.3900, 0.7084)$

Computation of confidence intervals

▶ 95% log-transformed confidence interval for the one year survival function:

$$\theta = \exp\left\{ \frac{Z_{1-\alpha/2}\sigma_S^2(t)}{\log[\widehat{S}(t_0)]} \right\} = \exp\left\{ \frac{1.96 \times 0.1479}{\log[0.5492]} \right\} = 0.6161$$

▶ CI: $(0.549211^{1/0.6165}, 0.5492^{0.6165}) = (0.3783, 0.6911)$

Computation of confidence intervals

▶ 95% arcsine-square root transformation confidence interval for the one year survival function

$$\sin^2\left\{\max\left[0, \arcsin(0.5492^{1/2}) - 0.5 \times 1.96 \times 0.1479 \times \left(\frac{0.5492}{1 - 0.5492}\right)^{1/2}\right]\right\}$$

to

$$\sin^2\left\{\min\left[\frac{\pi}{2}, \arcsin(0.5492^{1/2}) + 0.5 \times 1.96 \times 0.1479 \times \left(\frac{0.5492}{1 - 0.5492}\right)^{1/2}\right]\right\}$$

$$= (0.3903, 0.7032).$$

▶ three ways of computing pointwise confidence intervals

- ▶ how R compute the standard error for the survival?
- ▶ use bmt data in library KMsur

```
# estimate disease free survival for ALL group
> resALLDFS <- survfit(Surv(t2,d3) ~1, data=ALL)
> sresALLDFS <- summary(resALLDFS)
> sresALLDFS std.err
 [1] 0.02596722 0.03622354 0.04374408 0.04978449 0.05483610 0.05915279
 [7] 0.06288607 0.06613483 0.07143378 0.07357035 0.07540531 0.07696022
[13] 0.07825178 0.07952180 0.08050884 0.08122321 0.08167208 0.08185981
[19] 0.08178820 0.08145656 0.08086170 0.08026005 0.07929563
```

- ▶ R uses Greenwood's formula to estimate the variance of the product-limit estimator
- ▶ note that variance can be found with
  `summary.survfit(survival)`

Implement formula (4.2.2) Greenwood's variance for $S(t)$

```
> # number of events
> d<-sresALLDFS$n.event
> # number at risk
> y<-sresALLDFS$n.risk
> # estimate S(t) the product limit estimator
> S<- cumprod(1-d/y)
> # estimate the standard error with formula (4.2.2)
> varS<-S^2*cumsum(d/(y*(y-d)))
> # standard error for S(t)
> sqrt(varS)
 [1] 0.02596722 0.03622354 0.04374408 0.04978449 0.05483610 0.05915279
 [7] 0.06288607 0.06613483 0.07143378 0.07357035 0.07540531 0.07696022
[13] 0.07825178 0.07952180 0.08050884 0.08122321 0.08167208 0.08185981
[19] 0.08178820 0.08145656 0.08086170 0.08026005 0.07929563
```

- ▶ R uses Greenwood's formula to estimate the standard error of $S(t)$

```
> sresALLDFS std.err
 [1] 0.02596722 0.03622354 0.04374408 0.04978449 0.05483610 0.05915279
 [7] 0.06288607 0.06613483 0.07143378 0.07357035 0.07540531 0.07696022
[13] 0.07825178 0.07952180 0.08050884 0.08122321 0.08167208 0.08185981
[19] 0.08178820 0.08145656 0.08086170 0.08026005 0.07929563
```

- ▶ results from formula (4.2.2) implemented

```
> sqrt(varS)
 [1] 0.02596722 0.03622354 0.04374408 0.04978449 0.05483610 0.05915279
 [7] 0.06288607 0.06613483 0.07143378 0.07357035 0.07540531 0.07696022
[13] 0.07825178 0.07952180 0.08050884 0.08122321 0.08167208 0.08185981
[19] 0.08178820 0.08145656 0.08086170 0.08026005 0.07929563
```

- ▶ use library `km.ci` to look at different CI

```
> library(km.ci)
> library(help=km.ci)
> CI.ALLDFS <-km.ci(resALLDFS , conf.level=0.95, tl=NA, tu=NA,
 method="log")
# compare with results in sresALLDF
> unique(CI.ALLDFS$upper)
 [1] 1.0259327 1.0210933 1.0109066 0.9978320 0.9828316 0.9664027
 [7] 0.9488421 0.9303438 0.8910352 0.8703953 0.8491781 0.8274268
[13] 0.8051752 0.7819310 0.7581542 0.7338670 0.7090860 0.6838229
[19] 0.6580849 0.6318752 0.6051932 0.5770646 0.5483048
> sresALLDFS$upper
 [1] 1.0000000 1.0000000 1.0000000 0.9978320 0.9828316 0.9664027
 [7] 0.9488421 0.9303438 0.8910352 0.8703953 0.8491781 0.8274268
[13] 0.8051752 0.7819310 0.7581542 0.7338670 0.7090860 0.6838229
[19] 0.6580849 0.6318752 0.6051932 0.5770646 0.5483048
```

- ▶ implement formula (4.3.1) and (4.3.2) page 105 to look at the different values for CI

Comparison CI
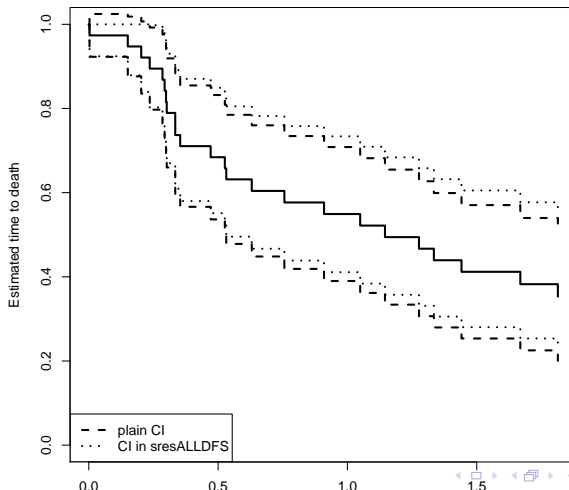
Outline

LU
MC

- ▶ Pointwise confidence intervals apply to a single point in the time scale
- ▶ look at simultaneous confidence bands (or confidence bands for short), which are valid for the entire range of time values simultaneously
- ▶ A 95% confidence band, for example, will capture the entire true survival curve about 19 out of 20 times.
- ▶ While the survival package doesn't offer tools for confidence bands, they may be calculated using `confBands` from the `OIsurv` library

```
data(bmt); attach(bmt)
my.surv <- Surv(t2[group==1], d3[group==1])
my.cb <- confBands(my.surv, confLevel=0.95, type="hall")
plot(survfit(my.surv ~ 1), xlim=c(100, 600), xlab="time",
        ylab="Estimated Survival Function",
        main="Confidence Bands for Example 4.2 in Klein/Moeschberger")

lines(my.cb$time, my.cb$lower, lty=3, type="s")
lines(my.cb$time, my.cb$upper, lty=3, type="s")
legend(100, 0.3, legend=c("K-M survival estimate",
      "pointwise intervals","confidence bands"), lty=1:3, bty="n")
 detach(bmt)
```

Reproducing Confidence Bands for Example 4.2 in Klein/Moeschberger
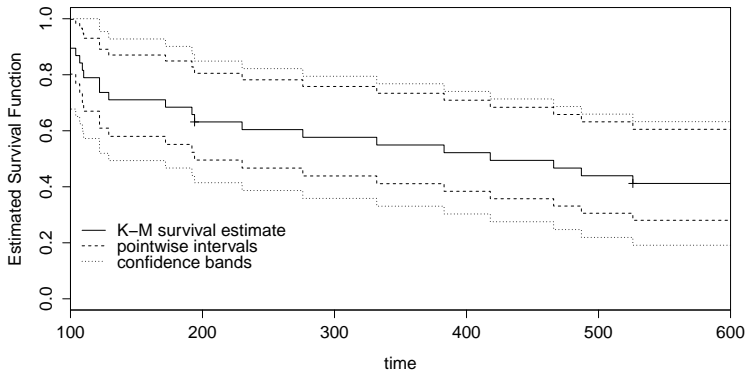
Outline

LU
MC

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

- the mean time to the event μ is given by $\mu = \int_0^\infty S(t)dt$

- estimator of μ obtained by substituting $\widehat{S}(t)$ for $S(t)$:
  $\hat{\mu} = \int_0^\infty \hat{S}(t)dt$ (compute the area under the KM curve)

- this estimator is appropriate only when the largest observation corresponds to a death

- in other cases, the product-limit estimator is not defined beyond the largest observation

- several solutions to this problem are available

- *solution 1*: changes the largest observed time to a death if it was a censored observation

- estimate the mean restricted to the interval 0 to t_{max}

- *solution 2*: estimate the meanlifetime restricted to some preassigned interval $[0, \tau]$
- τ is chosen by the investigator to be the longest possible time to which anyone could survive
- the estimated mean restricted to the interval $[0, \tau]$ in both cases is

$$\hat{\mu}_\tau = \int_0^\tau \widehat{S}(t)dt$$

Estimates of the variance and CI for the mean survival time

- variance for the mean survival time

$$\widehat{V}[\hat{\mu}_\tau] = \sum_{i=1}^{D} \left[\int_{t_i}^{\tau} \widehat{S}(t)dt \right]^2 \frac{d_i}{Y_i(Y_i - d_i)}$$

- confidence interval for the mean

$$\hat{\mu}_\tau \pm Z_{1-\alpha/2} \sqrt{\widehat{V}[\hat{\mu}_\tau]}$$

| KMsurv packages | CI S(t) | R code | Confidence intervals | Confidence bands | **Mean and median** $S(t)$ | Follow-up |
| 00000000 | 00000 | 000 | 0000000 | 000 | 000●0000 | 0000000000 |

Example

- ▶ Consider, for example, a sample of right censored life times $1, 1, 1^+, 2.5, 5^+, 7^+$
- ▶ The censored observation 7^+ will be treated as a 7 (true failure) for computing the Kaplan-Meier estimator so that $\hat{S}(7) = 0$
- ▶ Computations for KM

t_i	d_i	Y_i	$\hat{S}(t)$
1	2	6	2/3
2.5	1	3	4/9 $\frac{2}{3} \times \frac{2}{3}$
7	1	1	0 $\frac{2}{3} \times \frac{2}{3} \times 0$

KMsurv packages	CI S(t)	R code	Confidence intervals	Confidence bands	**Mean and median** $S(t)$	Follow-up
○○○○○○○○○	○○○○○	○○○	○○○○○○○○	○○○	○○○○●○○○	○○○○○○○○○○

Example

▶ estimated mean

$$\hat{\mu} = 1 \times (1 - 0) + \frac{2}{3} \times (2.5 - 1) + \frac{4}{9} \times (7 - 2.5) = 4$$

▶ in general: let $\tau_1 < \ldots < \tau_m$ be the distinct event (failure or censoring) times

$$\hat{\mu} = \sum_{i=1}^{n} \Delta\tau_i \hat{S}(\tau_{i-1})$$

▶ where $\tau_0 = 0$, $\Delta\tau_i = \tau_i - \tau_{i-1}$

KMsurv packages **CI S(t)** **R code** **Confidence intervals** **Confidence bands** **Mean and median** *S*(*t*) **Follow-up**
○○○○○○○○○ ○○○○○ ○○○ ○○○○○○○○ ○○○ ○○○○○●○○ ○○○○○○○○○○

Example

Median survival time

▶ the Product-Limit estimator can be used to provide estimates of quantiles of the distribution of the time-to-event distribution

0.5

▶ p*th* quantile: $x_p = \inf\{t : S(t) \leq 1 - p\}$ (x_p: smallest time at which the survival function is less than or equal to $1 - p$)

▶ $p = 1/2$; x_p is the median time to the event of interest

▶ estimate x_p: $\hat{x}_p = \inf\{t : \widehat{S}(t) \leq 1 - p\}$

▶ the standard error of x_p is difficult to compute because it requires an estimate of the density function of *X* at x_p

▶ The median survival time is the time at which half of the population has died and half are still alive

KMsurv packages **CI S(t)** **R code** **Confidence intervals** **Confidence bands** **Mean and median** $S(t)$ **Follow-up**
ooooooooo   ooooo ooo   ooooooooo   ooo   oooooo●o   oooooooooo

Example

- If $S_i(t_{Max}) > 0.5$: the median survival time cannot be estimated directly from the data
- there are methods to approximate the median survival time beyond the duration of follow-up based on geometric and linear growth in death rates
- The median and its 95% confidence interval may be estimated using `survfit()`

```
> data(drug6mp); attach(drug6mp)
> my.surv <- Surv(t1, rep(1, 21)) # all placebo patients observed
> survfit(my.surv ~ 1)
Call: survfit(formula = my.surv ~ 1)

records   n.max n.start  events  median 0.95LCL 0.95UCL
    21      21      21      21       8       4      12
```

KMsurv packages	**CI S(t)**	**R code**	**Confidence intervals**	**Confidence bands**	**Mean and median** $S(t)$	**Follow-up**
○○○○○○○○○	○○○○○ ○○○	○○○○○○○○	○○○	○○○○○○○●	○○○○○○○○○○

Example

▶ Using `survfit()` together with `print()`, the mean survival time and its standard error may be obtained:

```
> print(survfit(my.surv ~ 1), print.rmean=TRUE)
Call: survfit(formula = my.surv ~ 1)

   records      n.max     n.start      events     *rmean *se(rmean)
    21.00      21.00       21.00       21.00       8.67       1.38
    * restricted mean with upper limit =  23
> detach(drug6mp)
```

▶ The `print.rmean=TRUE` argument is used to obtain the mean and its standard error, and τ is automatically set as the largest observed or censored time. Alternatively, τ may be specified using the `rmean` argument

$$\begin{array}{c} L\ U \\ M\ C \end{array}$$

Outline

KMsurv packages

CI S(t)

R code

Confidence intervals

Confidence bands

Mean and median $S(t)$
　Example

Follow-up
　Descriptives
　Reverse Kaplan-Meier

KMsurv packages **CI S(t)** **R code** **Confidence intervals** **Confidence bands** **Mean and median** $S(t)$ **Follow-up**

Descriptives

Why follow-up?

▶ Why should we be bothered with follow-up?

▶ If we want to know 5-years survival after surgery for breast cancer for instance

▶ Would we be comfortable in reporting this if we have followed patients for three years?

▶ Follow-up tells you something about the maturity of the data, hence about the reliability of your results

▶ In reality, for instance in clinical trials, not all patients have been followed for the same length of time

▶ The first patient has been included 7 years ago

▶ The last patient has been included 2 years ago

▶ The follow-up is different from person to person

▶ So there is a follow-up *distribution*

KMsurv packages | **CI S(t)** | **R code** | **Confidence intervals** | **Confidence bands** | **Mean and median** $S(t)$ | **Follow-up**
oooooooo | ooooo | ooo | ooooooo | ooo | ooooooo | o●oooooooo

Descriptives

How to calculate median follow-up?

Wrong way

▶ Calculate the median of the follow-up times
▶ Why is this wrong?

follow-up time	status
1	0
2	0
3	1
4	1
5	0

3 is not a follow up time
because status = 1

KMsurv packages CI S(t) R code Confidence intervals Confidence bands Mean and median $S(t)$ **Follow-up**
○○○○○○○○○ ○○○○○ ○○○ ○○○○○○○○ ○○○ ○○○○○○○○ ○●○○○○○○○○

Descriptives

How to calculate median follow-up?

Wrong way

- ▶ Calculate the median of the follow-up times
- ▶ Why is this wrong?

Better

$$T = \min(X, C)$$

- ▶ Calculate the median of the follow-up times
- ▶ Exclude the patients that have died
- ▶ This is often done in practice

LU
MC

KMsurv packages | **CI S(t)** | **R code** | **Confidence intervals** | **Confidence bands** | **Mean and median** $S(t)$ | **Follow-up**
○○○○○○○○ | ○○○○○ | ○○○ | ○○○○○○○○ | ○○○ | ○○○○○○○○ | ○○●○○○○○○○○

Descriptives

Example

- ▶ Dutch Gastric Cancer Trial
- ▶ First patient died after 5 days; highest follow-up time: 11.27 years (alive)
- ▶ Removing dead patients: minimum follow-up 5.99 years, maximum 11.27 years, median follow-up 8.99 years

KMsurv packages ००००००००० **CI S(t)** ००००० **R code** ००० **Confidence intervals** ००००००००० **Confidence bands** ००० **Mean and median** $S(t)$ ००००००००० **Follow-up** ००००●००००००

Descriptives

Descriptives

After removing dead patients

Descriptives

			Statistic	Std. Error
survival since rando (yrs)	Mean		9,0051	,06591
	95% Confidence Interval for Mean	Lower Bound	8,8753	
		Upper Bound	9,1348	
	5% Trimmed Mean		9,0086	
	Median		8,9938	
	Variance		1,282	
	Std. Deviation		1,13209	
	Minimum		5,99	
	Maximum		11,27	
	Range		5,28	
	Interquartile Range		1,91	
	Skewness		,027	,142
	Kurtosis		-,857	,283

KMsurv packages CI S(t) R code Confidence intervals Confidence bands Mean and median $S(t)$ **Follow-up**
○○○○○○○○○ ○○○○○ ○○○ ○○○○○○○○ ○○○ ○○○○○○○○ ○○○○●○○○○○

Descriptives

Histogram

After removing dead patients

KMsurv packages   **CI S(t)**   **R code**   **Confidence intervals**   **Confidence bands**   **Mean and median** $S(t)$ **Follow-up**

Reverse Kaplan-Meier
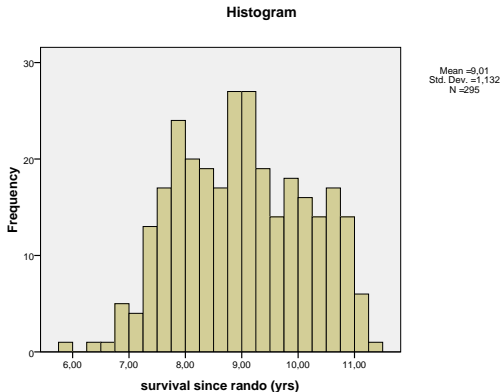
Reporting of follow-up

Often reported

▶ Median and range (minimum and maximum) follow-up

Less often reported

▶ The whole distribution

▶ Pity, really, but usually there is no space in a paper

▶ Both minimum and maximum don't tell you very much about maturity

▶ Especially minimum; there could be one patient who was lost to follow-up very early but last patient was included four years ago

KMsurv packages   **CI S(t)**   **R code**   **Confidence intervals**   **Confidence bands**   **Mean and median** $S(t)$ **Follow-up**
ooooooooo   ooooo   ooo   oooooooo   ooo   oooooooo   ooooooooooo

Reverse Kaplan-Meier

Estimating the censoring distribution

- ▶ How long it takes for the number at risk to drop to half the starting value?
- ▶ Two reasons for the number at risk to drop over time:
  - ▶ A subject can die or his data can be censored
  - ▶ If someone dies, you don't know how long they would have been followed
- ▶ From the point of view of tracking follow-up time, the roles of deaths and censoring are sort of reversed
- ▶ Idea: run the data through the Kaplan-Meier analysis again, but with the meaning of the status indicator reversed

KMsurv packages **CI S(t)** **R code** **Confidence intervals** **Confidence bands** **Mean and median** $S(t)$ **Follow-up**

Reverse Kaplan-Meier

Reverse Kaplan-Meier

- ▶ The end point is loss to follow-up (which is usually considered censoring)
- ▶ If a patient died, you can't know how long he would have been followed
- ▶ Death censors the true but unknown observation time of an individual
- ▶ Create a Kaplan-Meier curve where loss to follow-up is the event being followed, and a death is treated as censored data

LU
MC

KMsurv packages   **CI S(t)**   **R code**   **Confidence intervals**   **Confidence bands**   **Mean and median** $S(t)$ **Follow-up**

Reverse Kaplan-Meier

Follow-up using reverse Kaplan-Meier

One-minus-survival curve

One Minus Survival Function

KMsurv packages **CI S(t)** **R code** **Confidence intervals** **Confidence bands** **Mean and median** $S(t)$ **Follow-up**
00000000   00000   000   00000000   000   00000000   000000000●

Reverse Kaplan-Meier

Follow-up using reverse Kaplan-Meier

Mean and median table

Means and Medians for Survival Time

Mean[a]				Median		
		95% Confidence Interval				95% Confidence Interval
Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound
9,084	,064	8,958	9,210	9,068	,080	8,911

a. Estimation is limited to the largest survival time if it is censored.

Means and Medians for Survival Time

Median
95% Confidence Interval
Upper Bound
9,225

a. Estimation is limited to the largest survival time if it is censored.