

# University of Edinburgh, School of Mathematics

## Biostatistics (MATH11230), 2021/2022

Vanda Inácio

Here I reproduce the plots we have in the slides. They are easily obtained using the package `survival`. The crux is to create an object of class `Surv` containing the event times and whether or not they are censored.

```
require(survival)
times <- c(1, 3, 3, 6, 8, 9, 10)
status <- c(1, 1, 1, 0, 0, 1, 0)
S <- Surv(times, status)
S
```

```
## [1] 1 3 3 6+ 8+ 9 10+
```

```
class(S)
```

```
## [1] "Surv"
```

Note that when we print our object `S`, the censored times appear with a `+` symbol. We can then use the `survfit` function in order to obtain the Kaplan–Meier estimate of the survival function. Note that `conf.type = "plain"` implements the approach of slide 17 (based on Greenwood formula).

```
fit <- survfit(S ~ 1, conf.type = "plain")
fit
```

```
## Call: survfit(formula = S ~ 1, conf.type = "plain")
```

```
##
```

```
##      n  events  median 0.95LCL 0.95UCL
```

```
##      7      4      9      3      NA
```

```
summary(fit)
```

```
## Call: survfit(formula = S ~ 1, conf.type = "plain")
```

```
##
```

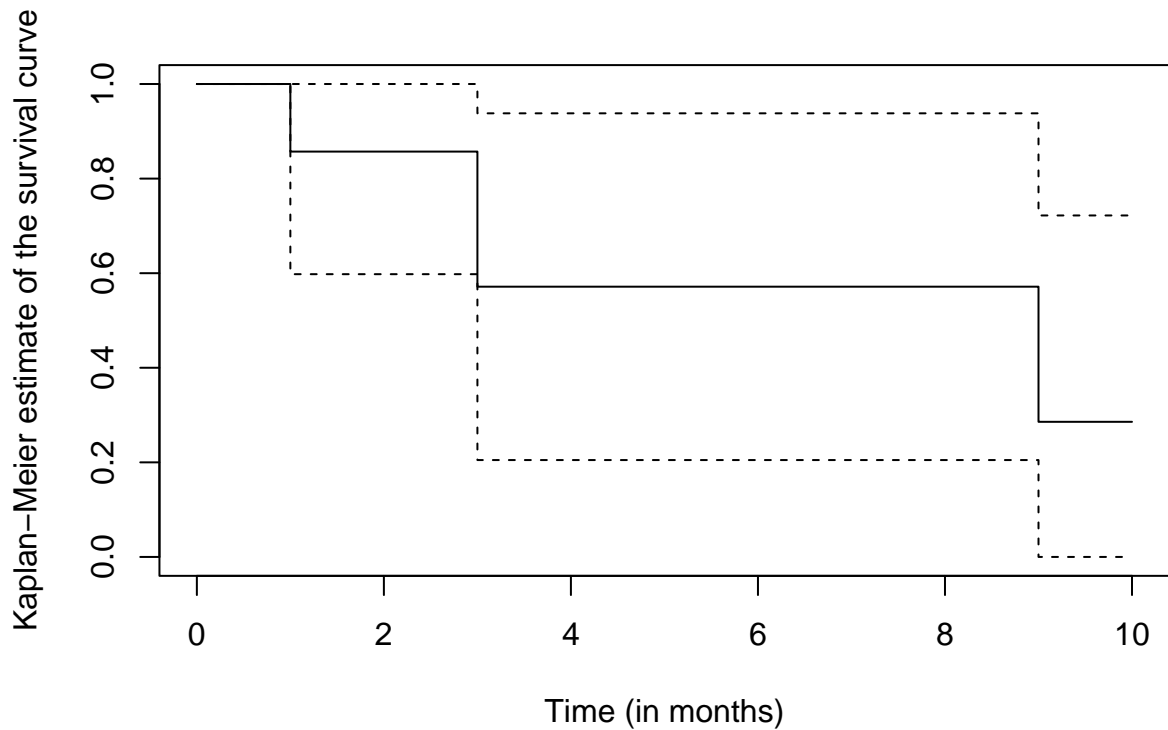
```
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
##    1     7      1   0.857   0.132    0.598    1.000
```

```
##    3     6      2   0.571   0.187    0.205    0.938
```

```
##    9     2      1   0.286   0.223    0.000    0.722
```

```
plot(fit,
      xlab = "Time (in months)",
      ylab = "Kaplan-Meier estimate of the survival curve")
```

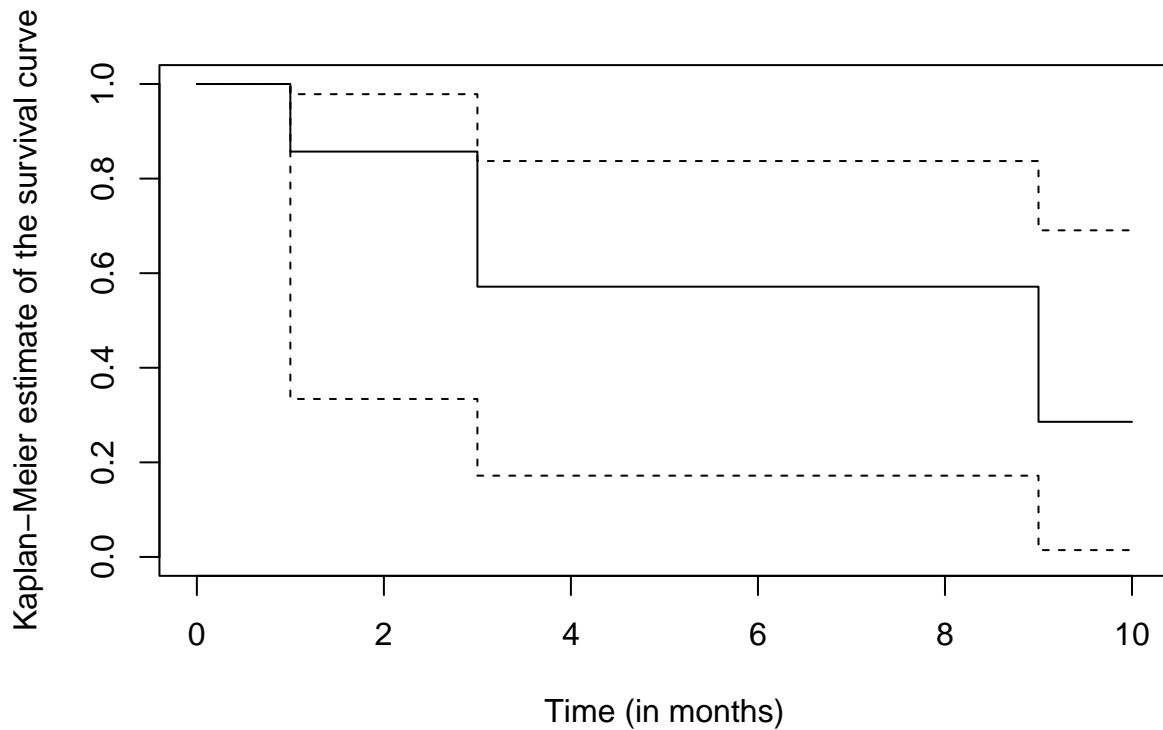


From the output we can see that the estimated median survival time is 9 (months). Other type of (pointwise) confidence intervals for the survival function can be obtained by changing the argument `conf.type`. A popular approach is to set it to `log-log`.

```
fit_alt <- survfit(S ~ 1, conf.type = "log-log")
summary(fit_alt)
```

```
## Call: survfit(formula = S ~ 1, conf.type = "log-log")
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1      7      1    0.857   0.132    0.3341    0.979
##    3      6      2    0.571   0.187    0.1719    0.837
##    9      2      1    0.286   0.223    0.0144    0.691
```

```
plot(fit_alt,
     xlab = "Time (in months)",
     ylab = "Kaplan-Meier estimate of the survival curve")
```



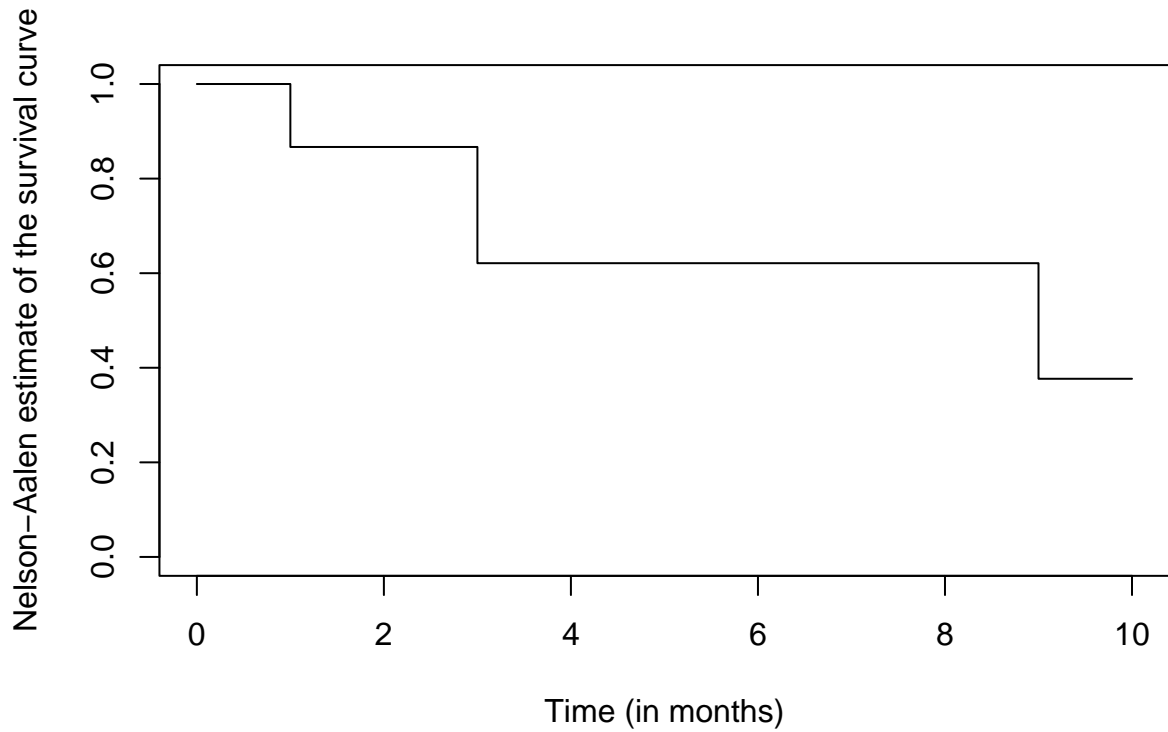
With regard to the Nelson–Aalen estimate, we proceed exactly in the same way as for the Kaplan–Meier estimate, we just need to specify the following arguments `stype = 2` and `ctype = 1`. More information can be find here:

<https://stat.ethz.ch/R-manual/R-devel/library/survival/html/survfit.formula.html>

```
fit_NA <- survfit(S ~ 1, stype = 2, ctype = 1)
summary(fit_NA)
```

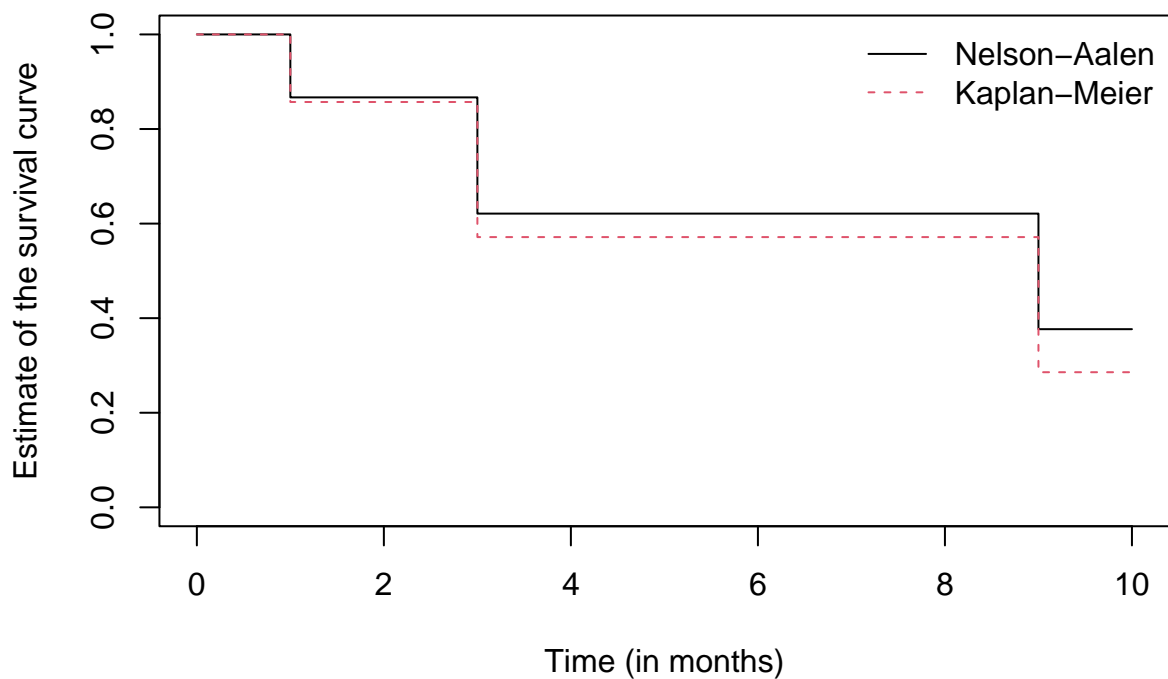
```
## Call: survfit(formula = S ~ 1, stype = 2, ctype = 1)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1      7      1   0.867   0.124    0.655      1
##    3      6      2   0.621   0.171    0.362      1
##    9      2      1   0.377   0.215    0.123      1
```

```
plot(fit_NA, conf.int = FALSE,
     xlab = "Time (in months)",
     ylab = "Nelson-Aalen estimate of the survival curve")
```



Let us visualise both the Kaplan-Meier and the Nelson-Aalen estimates in the same plot.

```
plot(fit_NA, conf.int = FALSE,
     xlab = "Time (in months)",
     ylab = "Estimate of the survival curve")
lines(fit, conf.int = FALSE, col = 2, lty = 2)
legend("topright", lty = c(1, 2), col = c(1, 2),
      c("Nelson-Aalen", "Kaplan-Meier"), bty = "n")
```



A recent package that produces very nice visualisations is `survminer`.

```
require(survminer)
toy_data <- data.frame("time" = times, "status" = status)
ggsurvplot(fit, conf.int = 0.95,
            ggtheme = theme_minimal(),
            data = toy_data)
```

