Kaplan-Meier estimate

SURVIVAL ANALYSIS IN R

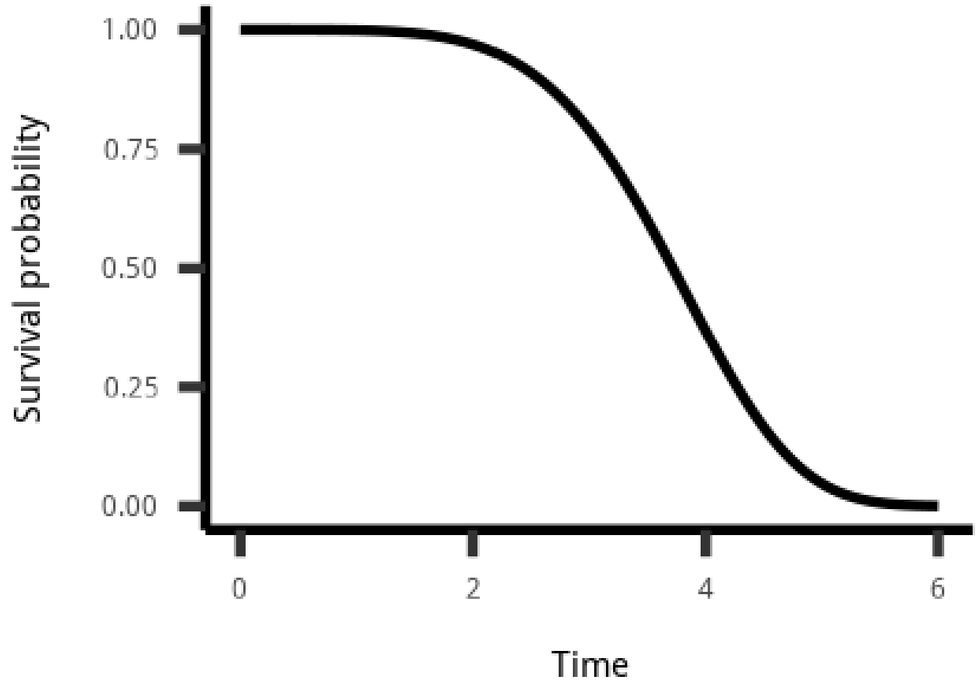
Heidi Seibold
Statistician at LMU Munich

datacamp

Survival function

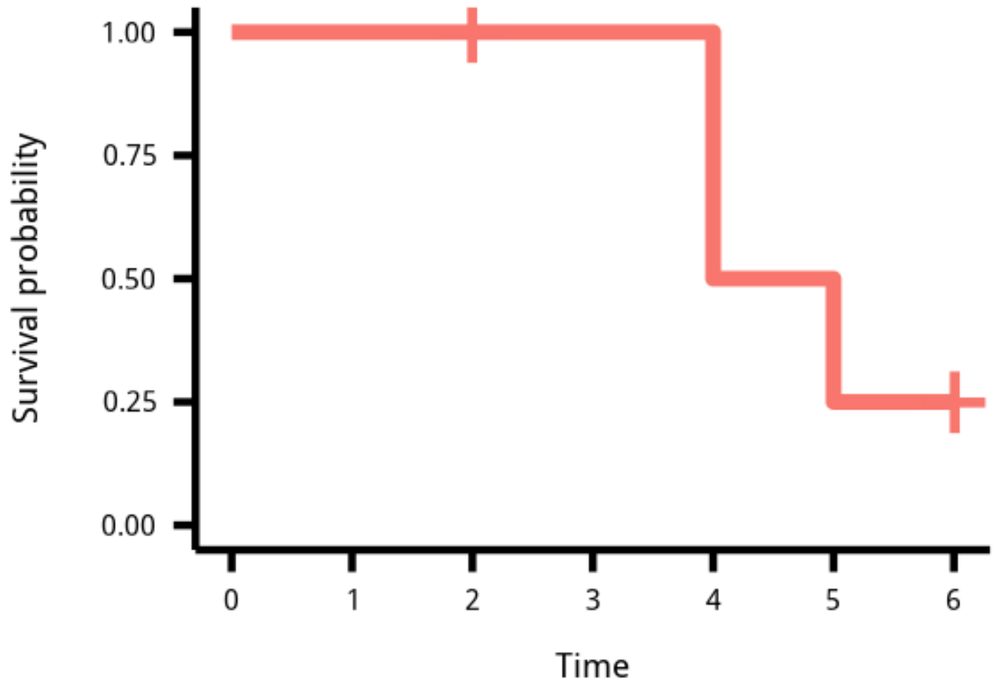
Theory

$$S(t) = 1 - F(t) = P(T > t)$$



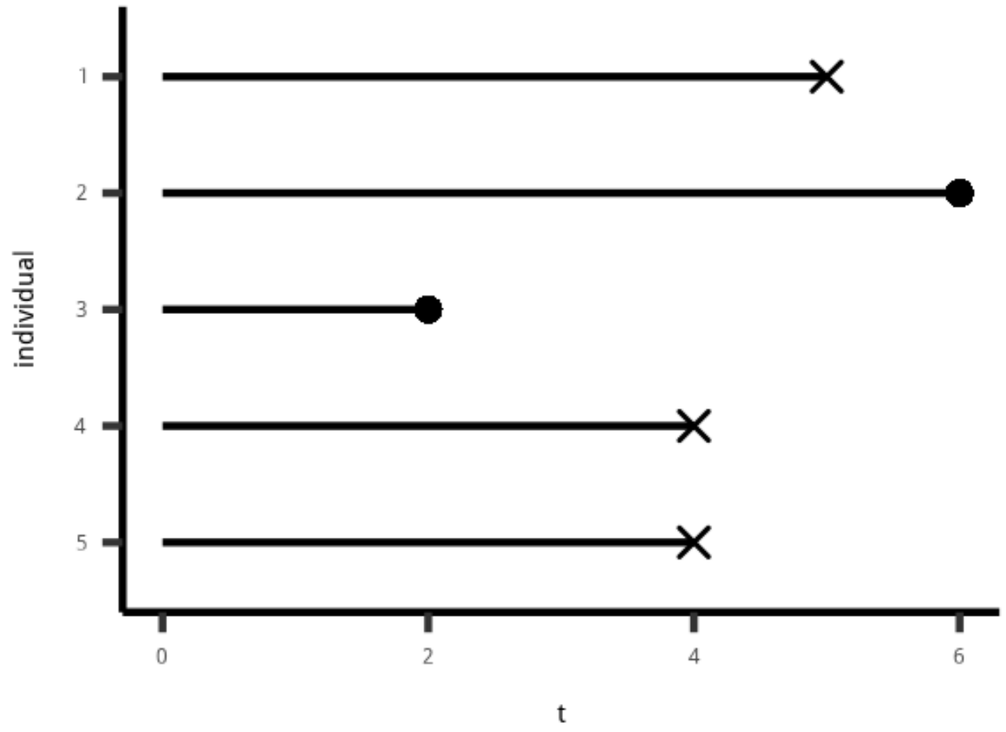
Estimation

$$\hat{S}(t) = \prod_{i:\, t_i \leq t} \frac{n_i - d_i}{n_i}$$



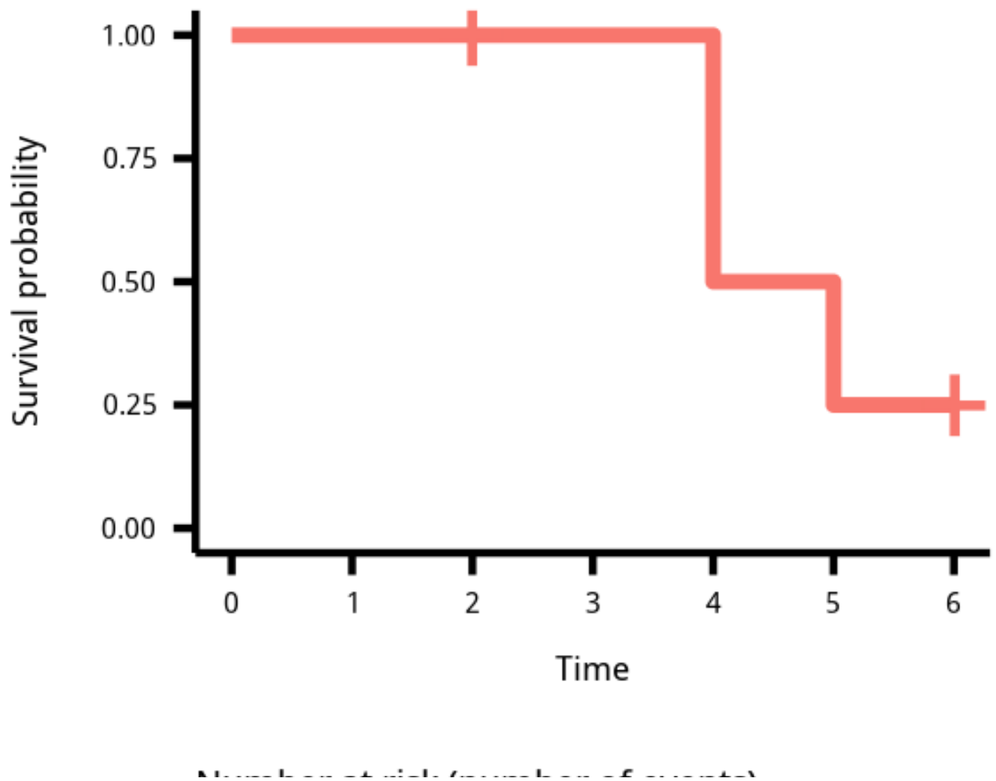
Survival function estimation

Data



Estimation

$$\hat{S}(t) = \prod_{i:\, t_i \leq t} \frac{n_i - d_i}{n_i}$$



Survival function estimation: Kaplan-Meier estimate

$$\hat{S}(t) = \prod_{i:\, t_i \leq t} \frac{n_i - d_i}{n_i}$$

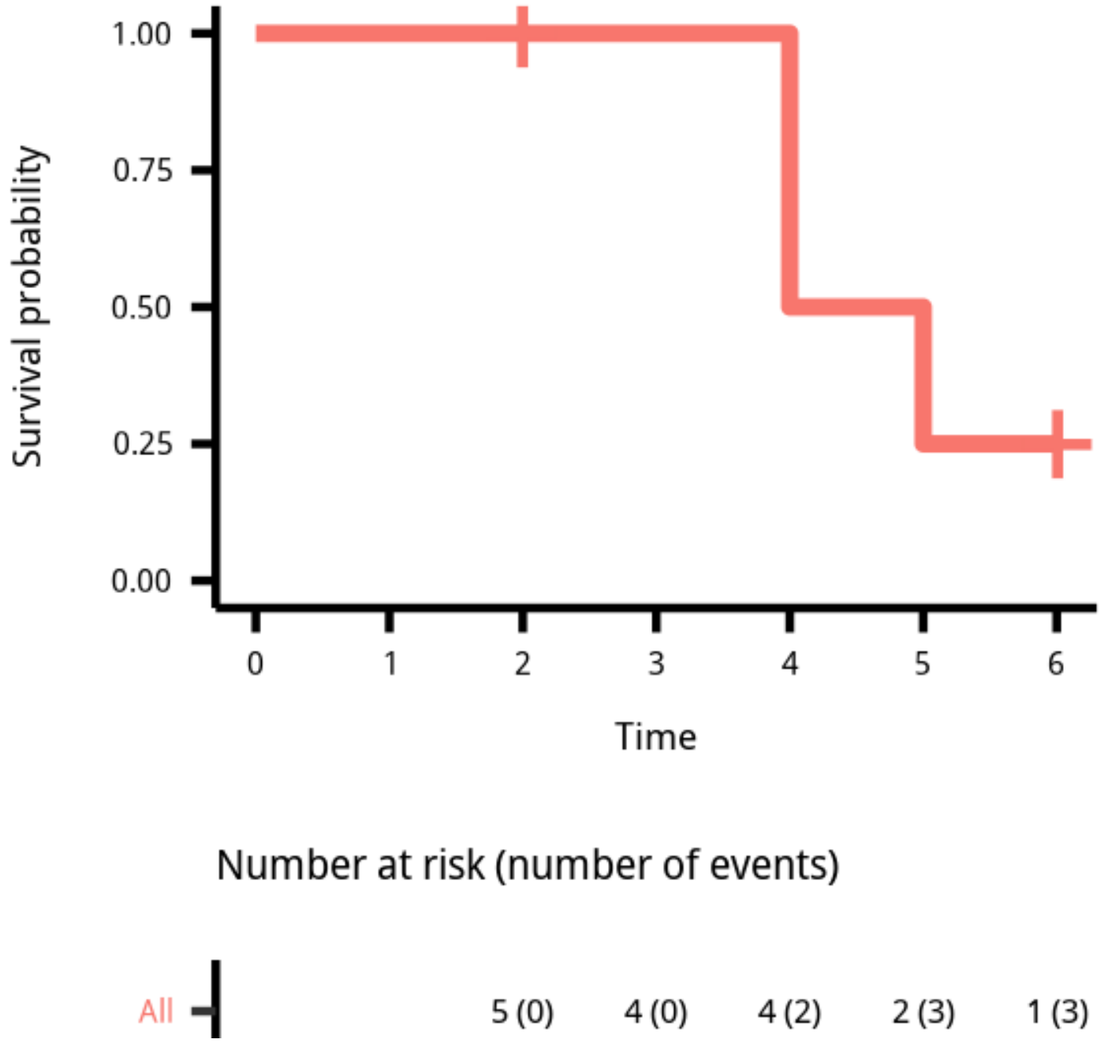
$$\hat{S}(2) = \frac{5-0}{5} = \frac{5}{5} = 1$$

$$\hat{S}(3) = \frac{4-0}{4} = \frac{4}{4} = 1$$

$$\hat{S}(4) = \frac{4-2}{4} = \frac{2}{4} = \frac{1}{2} = 0.5$$

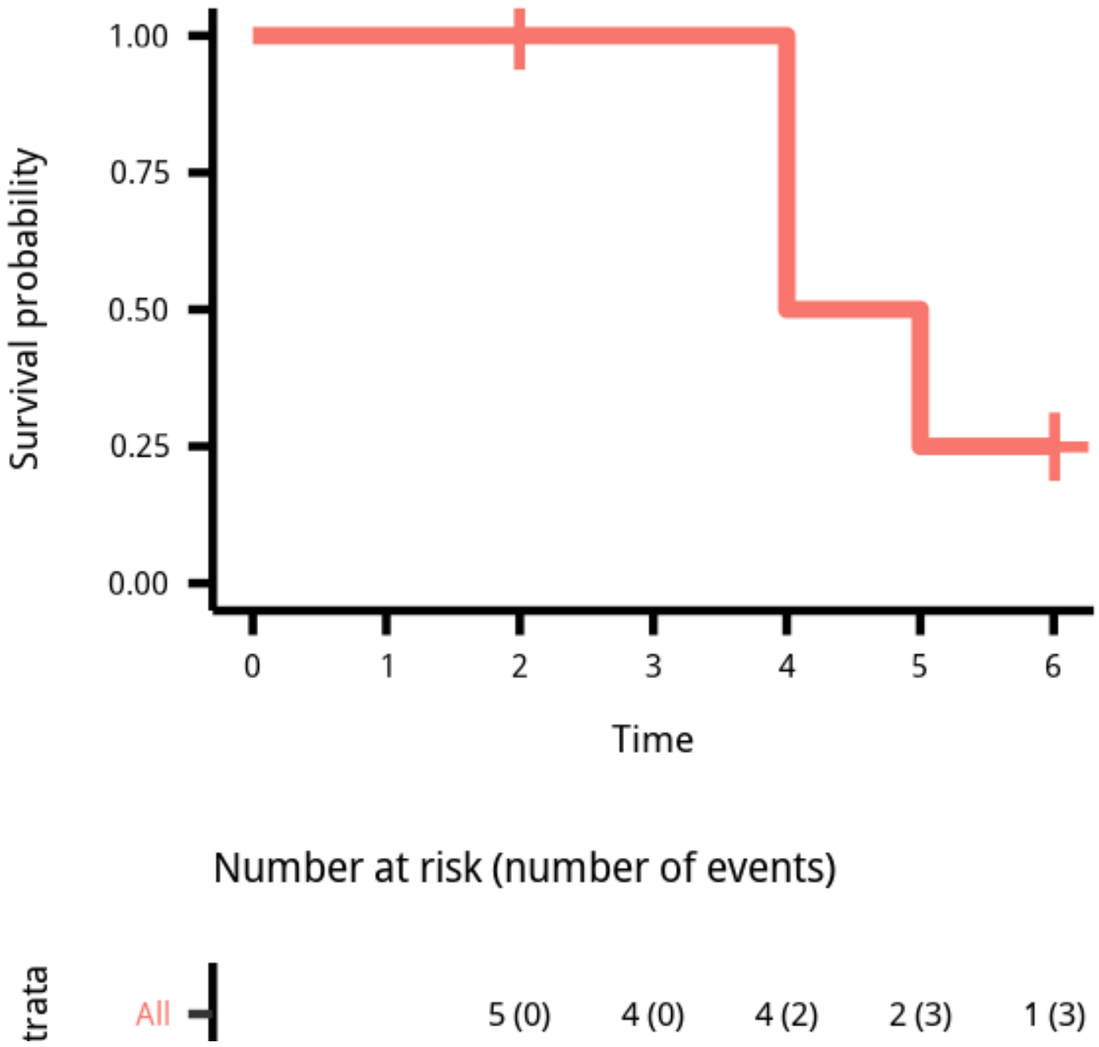
$$\hat{S}(5) = \frac{1}{2} \cdot \frac{2-1}{2} = \frac{1}{4} = 0.25$$

$$\hat{S}(6) = \frac{1}{4} \cdot \frac{1-0}{1} = \frac{1}{4} = 0.25$$



Survival function estimation: Kaplan-Meier estimate

```
km <- survfit(Surv(time, event) ~ 1)
ggsurvplot(km, conf.int = FALSE,
  risk.table = "nrisk_cumevents",
  legend = "none")
```



Let's practice!

SURVIVAL ANALYSIS IN R

datacamp

Understanding and visualizing Kaplan-Meier curves

SURVIVAL ANALYSIS IN R

Heidi Seibold
Statistician at LMU Munich

datacamp

The ggsurvplot function

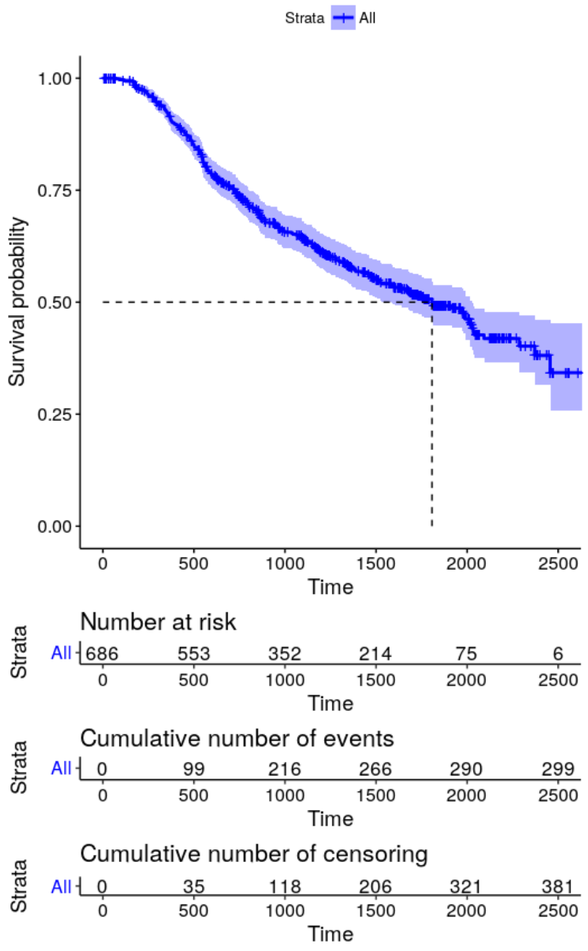
```
library(survminer)


ggsurvplot(fit)
```

```
ggsurvplot(
  fit, palette = NULL, linetype = 1,
  surv.median.line = "none", risk.table = FALSE,
  cumevents = FALSE, cumcensor = FALSE,
  tables.height = 0.25,
  ...
)
```

The ggsurvplot function

```
ggsurvplot(
  fit = km,
  palette = "blue",
  linetype = 1,
  surv.median.line = "hv",
  risk.table = TRUE,
  cumevents = TRUE,
  cumcensor = TRUE,
  tables.height = 0.1
)
```



The survfit function

```
survfit(object)
```

- If `object` is a `formula` : Kaplan-Meier estimation
- Other options for `object` (see upcoming chapters):
  - `coxph`
  - `survreg`

Let's practice!

SURVIVAL ANALYSIS IN R

datacamp

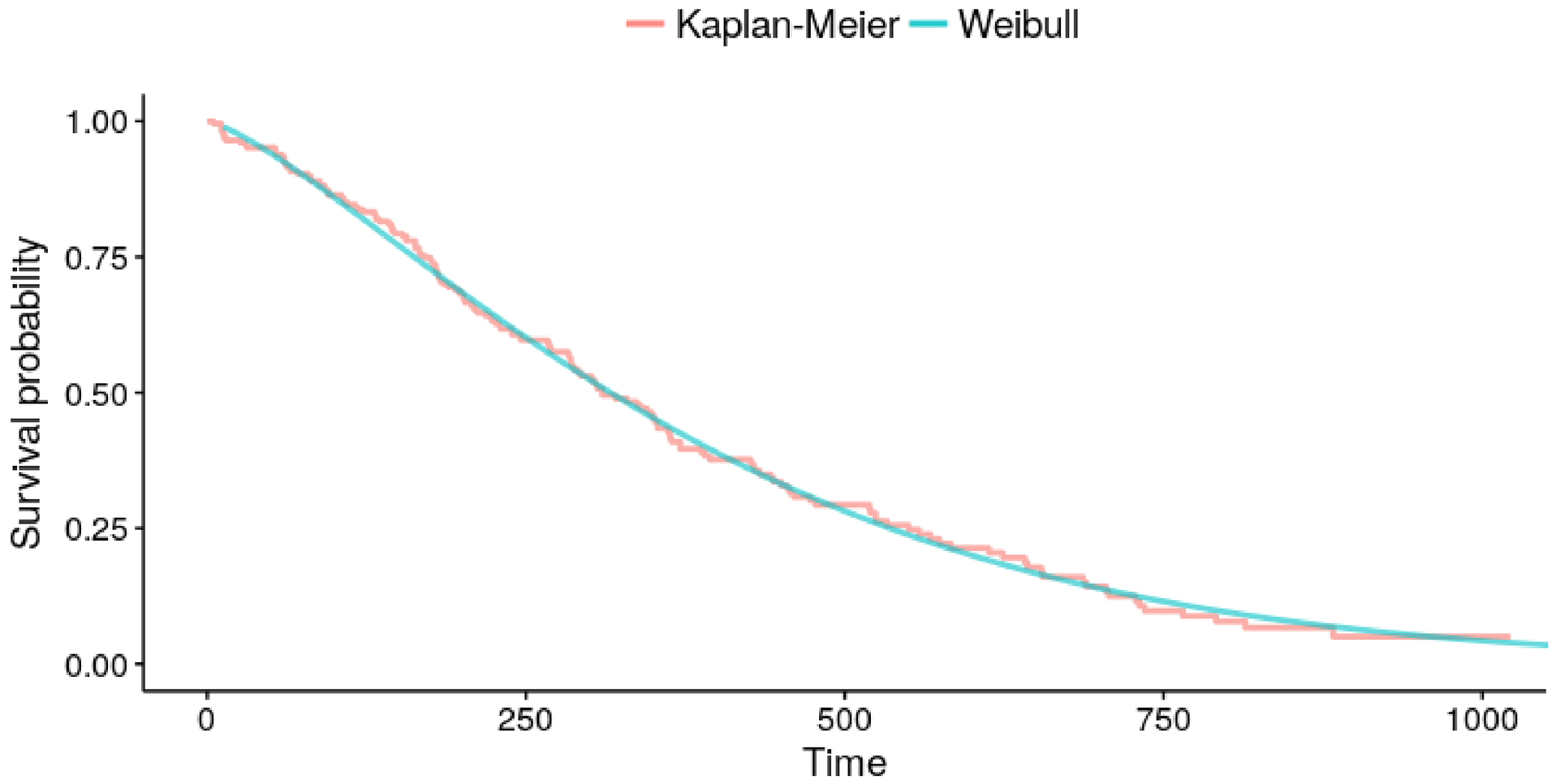
The Weibull model for estimating smooth survival curves

SURVIVAL ANALYSIS IN R

Heidi Seibold
Statistician at LMU Munich

datacamp

Why use a Weibull model?



Computing a Weibull model in R

Weibull model:

```
wb <- survreg(Surv(time, event) ~ 1, data)
```

Computing a Weibull model in R

Weibull model:

```
wb <- survreg(Surv(time, event) ~ 1, data)
```

Kaplan-Meier estimate:

```
km <- survfit(Surv(time, event) ~ 1, data)
```

Computing measures from a Weibull model

```
wb <- survreg(Surv(time, cens) ~ 1, data = GBSG2)
```

90 Percent of patients survive beyond time point:

```
predict(wb, type = "quantile", p = 1 - 0.9, newdata = data.frame(1))
```

```
       1
384.9947
```

`p = 1 - 0.9` because the distribution function is 1 - the survival function.

```
wb <- survreg(Surv(time, cens) ~ 1, data = GBSG2)
```

Survival curve:

```
surv <- seq(.99, .01, by = -.01)
t <- predict(wb, type = "quantile", p = 1 - surv, newdata = data.frame(1))
head(data.frame(time = t, surv = surv))
```

```
       time surv
1  60.6560 0.99
2 105.0392 0.98
3 145.0723 0.97
4 182.6430 0.96
5 218.5715 0.95
6 253.3125 0.94
```

Let's practice!

SURVIVAL ANALYSIS IN R

Visualizing the results of a Weibull model

SURVIVAL ANALYSIS IN R

Heidi Seibold
Statistician at LMU Munich

datacamp

Visualizing a Weibull model

Visualization tools often focus on step functions.

So the following code does NOT work:

```
wb <- survreg(Surv(time, cens) ~ 1)
ggsurvplot(wb)
```

Visualizing a Weibull model

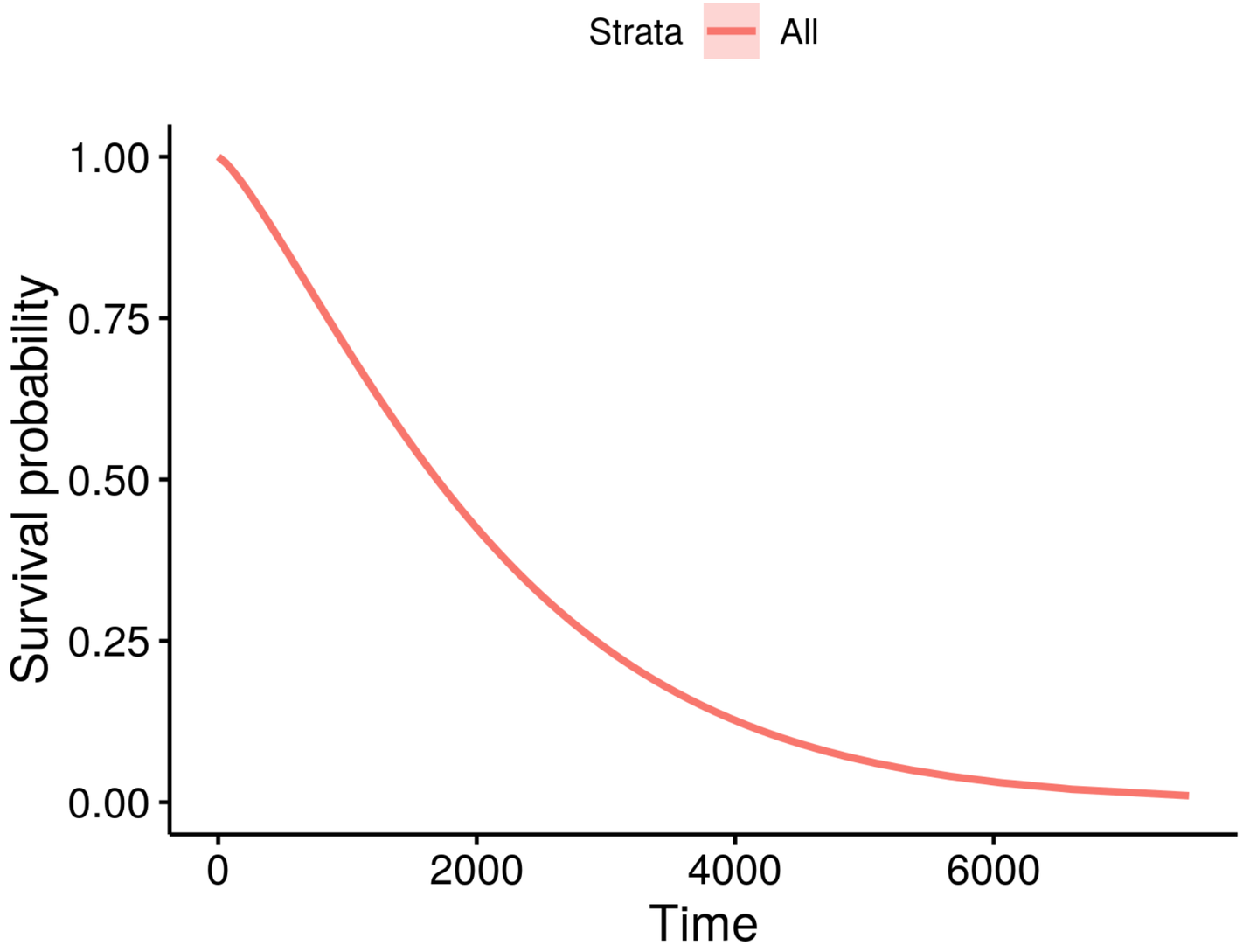
```
wb <- survreg(Surv(time, cens) ~ 1)
```

Survival curve:

```
surv <- seq(.99, .01, by = -.01)
t <- predict(wb, type = "quantile", p = 1 - surv, newdata = data.frame(1))
surv_wb <- data.frame(time = t, surv = surv,
                      upper = NA, lower = NA, std.err = NA)
```

Plot:

```
ggsurvplot_df(fit = surv_wb, surv.geom = geom_line)
```


Strata
All
1.00
0.75
0.50
0.25
0.00
Survival probability
0
2000
4000
6000
Time

Let's practice!

SURVIVAL ANALYSIS IN R

datacamp