Biostatistics

Survival analysis

Burkhardt Seifert & Alois Tschopp

Biostatistics Unit
University of Zurich

Survival analysis

- Data about life times ("time to event")

  Examples:
  - Time between start of therapy and death (survival time)
  - Time between discharge and re-hospitalisation
  - Time between surgery and relapse

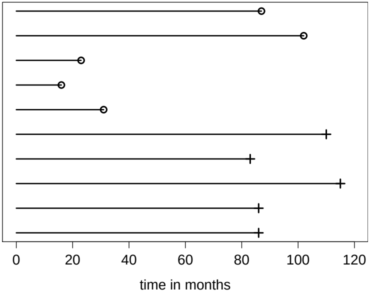
- Fundamentally new: data are censored

  Examples:
  - End of study before death of all patients
  - Other cause of death

In general: Only partly it is known which minimal time a patient has survived.

Survival function

Example: Survival time of $n = 116$ patients with melanoma stage 1 after surgery:

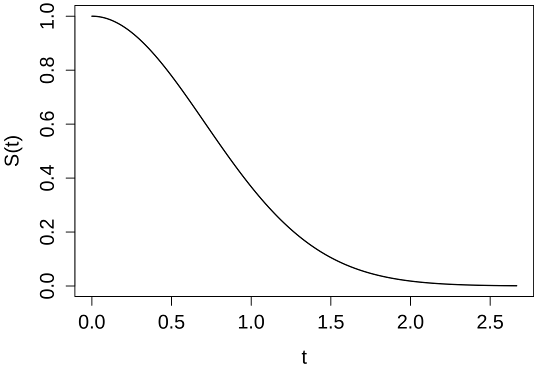


nr	sex	Breslow	time	state
2	1	2.00	87	1
3	2	2.13	102	1
4	1	2.07	23	1
6	2	2.23	16	1
7	1	1.87	31	1
22	1	3.40	110	0
23	1	2.68	83	0
24	1	0.46	115	0
28	1	1.11	86	0
30	2	0.64	86	0

Survival function

Distribution function of the survival times: $F(t)$

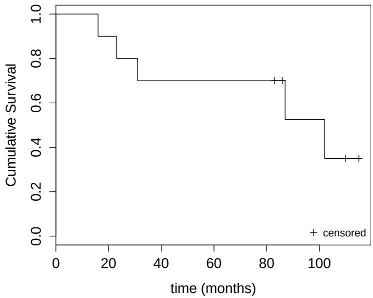
Survival function: $S(t) = 1 - F(t)$



Proportion of patients who survive a certain time point.

Kaplan-Meier estimator of the survival function

Idea: conditional probabilities (law of total probability)



p_1 = probability to survive day 1

p_2 = probability to survive day 2, given one has survived day 1

$$\hat{S}(0) = 1$$

$$\hat{S}(1) = \hat{p}_1$$

$$\hat{S}(2) = \hat{S}(1) \times \hat{p}_2$$

$$\dots$$

$$\hat{S}(t) = \hat{S}(t-1) \times \hat{p}_t$$

Kaplan-Meier estimator of the survival function

n_t = number of patients at time t

- Somebody is dying

$$\hat{p}_t = \frac{n_t - 1}{n_t}, \quad n_{t+1} = n_t - 1$$

- Somebody is censored

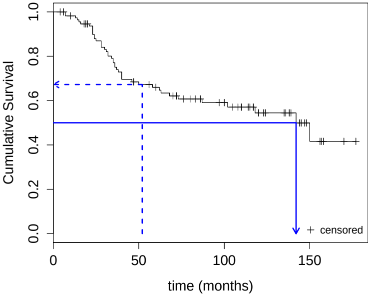
$$\hat{p}_t = \frac{n_t - 0}{n_t}, \quad n_{t+1} = n_t - 1$$

$$\boxed{\hat{S}(t) = \hat{S}(t-1) \times \hat{p}_t = \hat{p}_1 \times \hat{p}_2 \times \ldots \times \hat{p}_t}$$

```
Call: survfit(formula = Surv(time, state) ~ 1, data = melanom)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   16     10       1    0.900  0.0949        0.732            1
   23      9       1    0.800  0.1265        0.587            1
   31      8       1    0.700  0.1449        0.467            1
   87      4       1    0.525  0.1865        0.262            1
  102      3       1    0.350  0.1894        0.121            1
```

Description of survival times



Reasonable information:

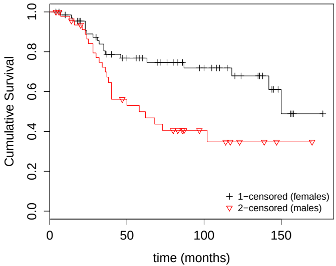
- mean and median-follow-up (descriptive)
- median survival time from K-M curve (if possible)
- survival probability from K-M curve after 1, 5, 10, . . . years ± standard error
- patients under risk after 1, 2, . . ., 10, . . . years.

Pointless information:

- mean survival time from K-M curve
- mean survival time of the deceased (descriptive)
- proportion (%) of the deceased

Comparison of survival functions

Example: Comparison of sex-specific melanoma survival



Scientific hypothesis:
Females (sex=1) have a different survival function than males (sex=2).

Comparison of two or more survival functions: Logrank-test

```
Call:
survdiff(formula = Surv(zeit.total, status) ~ sex, data = melanom)

       N Observed Expected (O-E)^2/E (O-E)^2/V
sex=1 68       19     27.6      2.70       7.8
sex=2 48       24     15.4      4.87       7.8

 Chisq= 7.8  on 1 degrees of freedom, p= 0.00524
```

$\Rightarrow$ Significant gender-difference. (A χ^2-test is wrong here!)

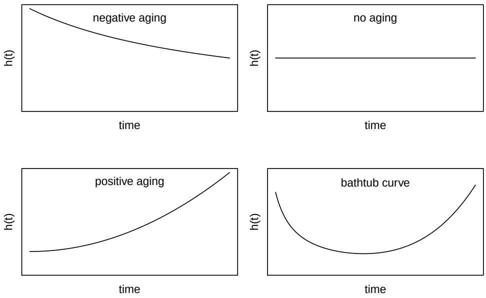
Hazard function

("force of mortality"):

Definition:

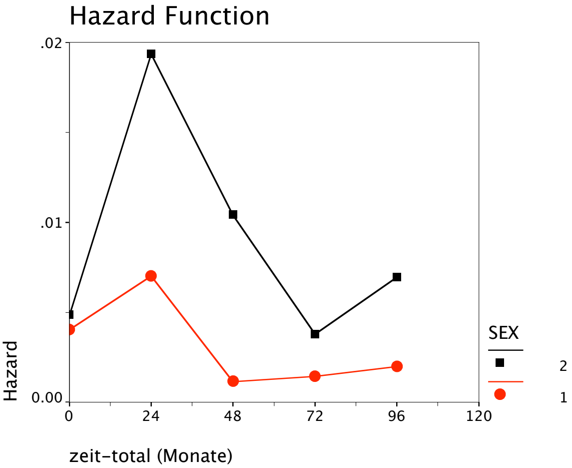
$$h(t) = f(t)/S(t) \quad \text{where: } f = F'$$

$h(t)$ = event rate at time t conditional on survival until time t or later.



Hazard function

Example: Comparison of gender-specific hazard functions



Cox-regression

("proportional hazards model"):

Question: How can we investigate the influence of $x_1, \ldots, x_k$ on the survival function $S(t) = P(y = 0|t)$?

Modelling by means of the hazard function:

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$

$$h_0(t) = \text{"baseline hazard"}$$

Relative hazard when increasing x_1 by one unit:

$$\frac{h(t|x_1 = a+1)}{h(t|x_1 = a)} = \frac{h_0(t) \times \exp(\beta_1 \times (a+1) + \ldots)}{h_0(t) \times \exp(\beta_1 \times (a) + \ldots)} = \exp(\beta_1)$$

analogous to odds ratio in logistic regression.

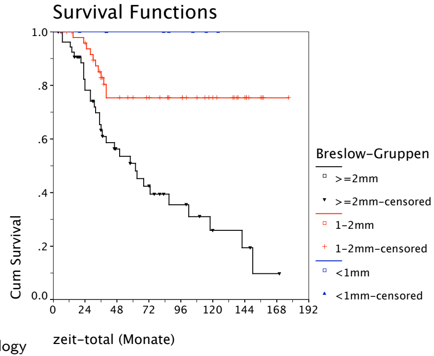
Cox-regression

Example: Melanoma, x = Depth of the tumor (logarithmised)

```
Call:
coxph(formula = Surv(zeit.total, status) ~ ln.Breslow., data = melanom)


             coef exp(coef) se(coef)    z       p
ln.Breslow. 1.27      3.57     0.22 5.79 6.9e-09

Likelihood ratio test=34.5  on 1 df, p=4.33e-09  n=115
(1 observation deleted due to missingness)
```



Literature

Hosmer, D. W., Lemeshow S., and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley, 2nd edition, 392 pages.

Examples worked out for `Stata`, `SPSS`, `SAS`, (`R`) at `http://www.ats.ucla.edu/stat/examples/asa2/default.htm`

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, 502 pages.

Examples worked out for `SAS` at `http://www.ats.ucla.edu/stat/examples/sakm/`

Chapters in:

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman & Hall.

Armitage, P., Berry, G., and Matthews, J. N. S. (2002). *Statistical methods in medical research*. Blackwell, 4th edition.

Hüsler, J. and Zimmermann, H. (2006). *Statistische Prinzipien für medizinische Projekte*. Hans Huber, Bern.
- clearly presented textbook, 355 pages.

Kirkwood, B. R. and Sterne, J. A. C. (2006). *Essential Medical Statistics*. Blackwell, 4th edition.

Matthews, D. E. and Farewell, V. T. (1988). *Using and understanding medical statistics*. Karger, 2nd edition.