

# L24: Survival analysis

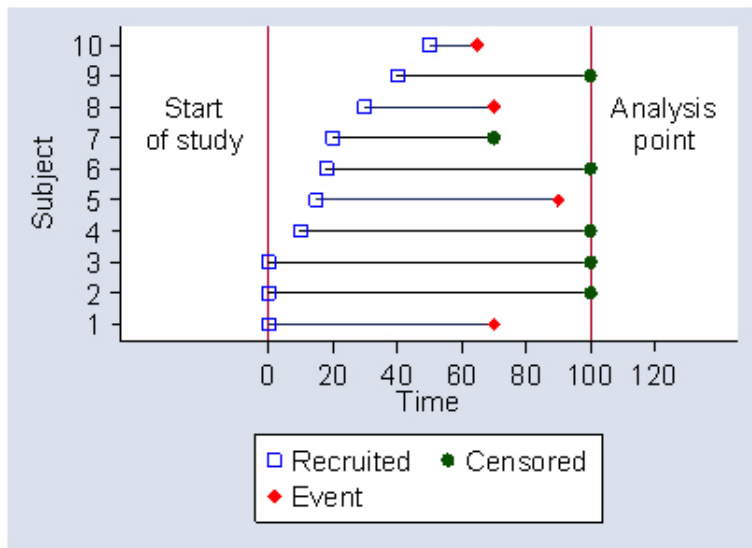
Department of Statistics  
University of South Carolina

Elementary Statistics for the Biological and Life Sciences  
(STAT 205)

# Survival-times after cardiac allografts

- Four authors: Messmer BJ, Nora JJ, Leachman RD, and Cooley DA.
- The Lancet. May 10, 1969
- 57 patients who were eligible for heart transplants
- Patients were divided into those who did and did not receive a transplant
- Either time to death or follow-up time was recorded
- Mean time to death or follow-up time was recorded
- The mean was higher in the group who received transplant (111 days versus 74 days)

# What does the data look like?



# The censoring problem

- Not every subject died at the end of experiment period (analysis point).
- We know the exact survival time (death time/event time) for each subject with a “red dot” mark in the plot.
- For the rest who have “green dot” marks, we only know they were alive at the end of experiment period (or missing point), however, we do not know when would they die in the future. This is called **right-censored survival data**.
- By utilizing the follow-up times for the patients who were still alive in the data analysis, statistically speaking, what have the authors done to these patients?

# Survival-times after cardiac allografts

The authors treated the living patients as if they died on their last day of follow-up. Statistically speaking they murdered the patients still alive!



What statistical procedure correctly accounts for patients who are lost to follow-up or still alive (censored)?

# Survival Analysis

- Analyzes time-to-event data.
  - Example of events: include death, recurrence of disease, or device failure
- Can answer questions such as:
  - What proportion of the population will survive past a certain time?
  - Of those that survive, at what rate will they experience the event?
  - Is there a significant difference in the time to an event between two groups?

- Censoring occurs when incomplete information is available about the survival time of some subjects.
- Censoring can occur when a subject:
  - Experiences the event before study begins but time is unknown (left censoring);
  - Does not experience event before study ends (right censoring);
  - Does not complete study (e.g., drops out or is lost to follow up) (random censoring).

# Kaplan-Meier formula

Let

$n_i$ : be the number of patients known to be **at risk** at the beginning of day  $i$

$d_i$ : be the number of patients who **die** on day  $i$ .

Then for patients alive at the beginning of the  $i$ -th day, the estimated probability of surviving the day is:

$$p_i = \frac{n_i - d_i}{n_i}$$

The probability that a patient survives the first  $t$  days is the product probability of surviving days 1, 2, ...,  $t$  which is estimated by

$$S(\hat{t}) = \prod_{\{k:t_k < t\}} p_k$$



## Example 1

Consider a clinical trial in which 8 patients have the following survival pattern (in months): 2, 2+, 3, 4, 4+, 5, 6, 7 (2+ and 4+ are individuals lost to follow-up at 2 and 4 months, respectively).

$$\hat{S}(t = 2) = (8 - 1)/8 = 0.875$$

$$\hat{S}(t = 3) = (0.875) \times [(6 - 1)/6] = 0.729$$

$$\hat{S}(t = 4) = 0.875 \times 0.729 \times [(5 - 1)/5] = 0.583$$

$$\hat{S}(t = 5) = 0.875 \times 0.729 \times 0.583 \times [(3 - 1)/3] = 0.389$$

$$\hat{S}(t = 6) = 0.875 \times 0.729 \times 0.583 \times 0.389 \times [(2 - 1)/2] = 0.194$$

$$\hat{S}(t = 7) = 0$$

Gross AJ and Clark VA. 1975. Survival distributions: reliability applications in the biomedical sciences. P. 46.

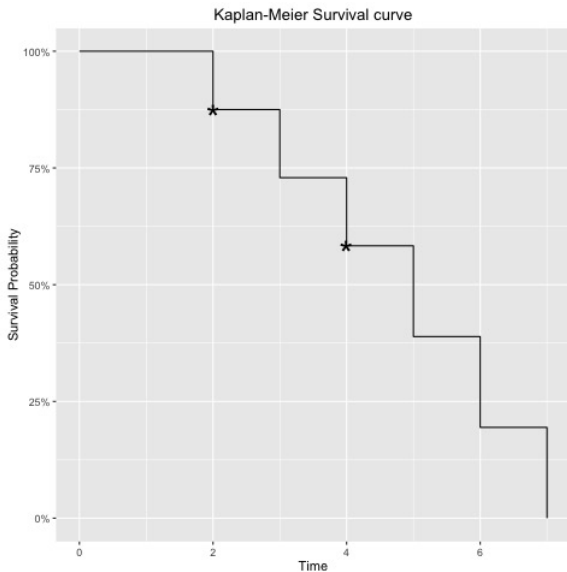
# R code for example 1

```
install.packages("survival")
install.packages("ggfortify")
library(survival)
library(ggfortify)
time<-c(2, 2, 3, 4, 4, 5, 6, 7)
censor<-c(1, 0, 1, 1, 0, 1, 1, 1)
fit<-survfit(Surv(time, censor) ~1)
summary(fit)
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
2	8	1	0.875	0.117	0.6734	1
3	6	1	0.729	0.165	0.4680	1
4	5	1	0.583	0.186	0.3127	1
5	3	1	0.389	0.201	0.1410	1
6	2	1	0.194	0.170	0.0349	1
7	1	1	0.000	NaN	NA	NA

```
autoplot(fit, censor.shape="*", censor.size=10, conf.int=F,
  ylab="Survival Probability", xlab="Time",
  main="Kaplan-Meier Survival curve")
```

# Example 1 Survival Plot



## Example 2

Example 2: Consider a clinical trial in which 12 patients have the following survival pattern (in months): 6, 8, 8+, 10, 10, 10+, 12, 16, 16, 16, 20, 26 (8+ and 10+ are individuals lost to follow-up at 8 and 10 months, respectively).

$$\hat{S}(t = 6) = (12 - 1)/12 = 0.917$$

$$\hat{S}(t = 8) = (0.917) \times [(11 - 1)/11] = 0.834$$

$$\hat{S}(t = 10) = 0.917 \times 0.824 \times [(9 - 2)/9] = 0.649$$

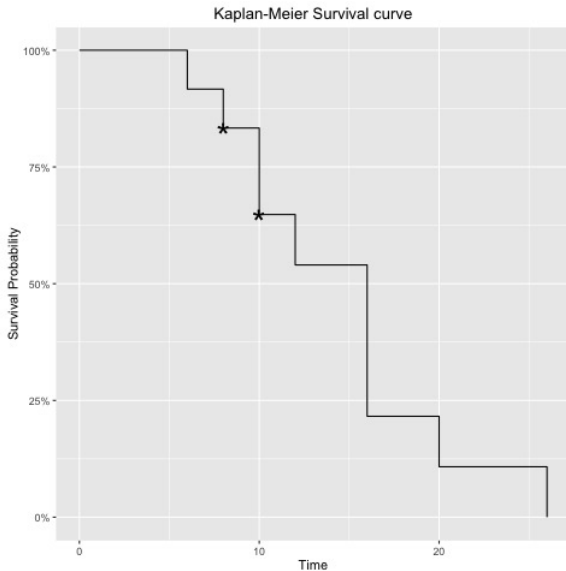
$$\hat{S}(t = 12) = 0.917 \times 0.824 \times 0.649 \times [(6 - 1)/6] = 0.541$$

$$\hat{S}(t = 16) = 0.917 \times 0.824 \times 0.649 \times 0.541 \times [(5 - 3)/5] = 0.216$$

$$\begin{aligned}\hat{S}(t = 20) &= 0.917 \times 0.824 \times 0.649 \times 0.541 \times 0.216 \times [(2 - 1)/2] \\ &= 0.108\end{aligned}$$

$$\hat{S}(t = 26) = 0$$

# Example 2



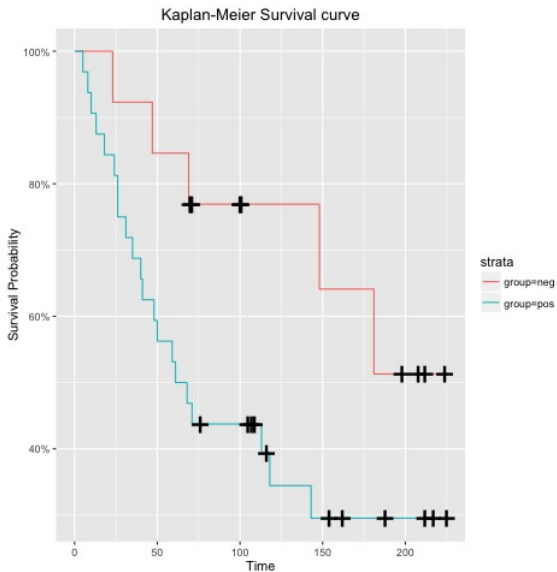
# HPA staining and breast cancer

- We consider data from a retrospective study of 45 women who had surgery for breast cancer. Tumor cells, surgically removed from each woman, were classified according to the results of staining on a marker taken from the Roman snail, the *Helix pomatia* agglutinin (HPA).
- The **survival times** in months  $t_i$  and staining results ( $x_i = 0$  for negative and  $x_i = 1$  for positive) for the 45 women are given. Censoring status was also reported.

# The survival time data

- $x = 0$  staining:  
23; 47; 69; 70\*; 71\*; 100\*; 101\*; 148; 181; 198\*; 208\*; 212\*; 224\*
- $x = 1$  staining:  
5; 8; 10; 13; 18; 24; 26; 26; 31; 35; 40; 41; 48; 50; 59; 61; 68;  
71; 76\*; 105\*; 107\*; 109\*; 113; 116\*; 118; 143; 154\*; 162\*; 188\*;  
212\*; 217\*; 225\*
- \* indicates right-censoring.
- What is the estimate of survival function for each group?

# Comparison of two survival plots





- Compares survival curves among independent groups.
  - Example: Comparison of survival times in cancer patients who are given a new treatment with patients who receive standard chemotherapy.
- $H_0$  : there is no difference in the population survival curves of the groups.
- $H_A$  : difference exists in the population survival curves of the groups.
- Adjusts for censoring.

# Comparison test in R

```
install.packages("survival")
install.packages("ggfortify")

library(survival)
library(ggfortify)
timeneg <- c(23,47,69,70,71,100,101,148,181,198,208,212,224)
censneg <- c(1,1,1,0,0,0,0,1,1,0,0,0,0)
timepos <- c(5, 8, 10, 13, 18, 24, 26, 26, 31, 35, 40, 41, 48,
            50, 59, 61, 68, 71, 76, 105, 107, 109, 113, 116,
            118, 143, 154, 162, 188, 212, 217, 225)
censpos <- c(rep(1,times=18),0,0,0,0,1,0,1,1,0,0,0,0,0,0)

group <- as.factor( c( rep("neg",times=13), rep("pos",times=32) ) )
time <- c(timeneg,timepos)
cens <- c(censneg,censpos)
fit <- survfit(Surv(time, cens) ~ group)
summary(fit)
autoplot(fit, conf.int=F, censor.shape="+", censor.size=10,
         ylab="Survival Probability", xlab="Time",
         main="Kaplan-Meier Survival curve")
survdiff( Surv(time,cens) ~ group)
```

# Comparison test in R

Call:

```
survdifff(formula = Surv(time, cens) ~ group)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
group=neg	13	5	9.57	2.18	3.51
group=pos	32	21	16.43	1.27	3.51

Chisq= 3.5 on 1 degrees of freedom, p= 0.0608

p-value 0.0608 is greater than 5% level of significance, we accept the null and claim that the survival curve for positive and negative groups are the same.