Kaplan-Meier Estimation Method

Dr Rahimzadeh, Biostatistics

Goals of Survival Analysis

- **Goal 1:** To estimate and interpret survivor and/ or hazard functions from survival data
- **Goal 2:** To compare survivor and/or hazard functions.
- **Goal 3:** To assess the relationship of explanatory variables to survival time

Notation

- We have a sample of failure times $\Rightarrow$ What is the available information?

- Notation (i denotes the patient)
  - $\triangleright$ T_i^* 'true' time-to-event
  - $\triangleright$ because of censoring we do not always observe T_i^*
  - $\triangleright$ C_i the censoring time (e.g., the end of the study or a random censoring time)

- Available data for each patient
  - $\triangleright$ observed event time: $T_i = \min(T_i^*, C_i)$
  - $\triangleright$ event indicator: $\delta_i = 1$ if event; $\delta_i = 0$ if censored

Notation

Patient	T_i^*	C_i	T_i	δ_i
1	3.5	—	3.5	1
2	3.4	2.2	2.2	0
3	5.7	5	5	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

The end of the study is at 5 years

Assumptions of Kaplan-Meier Estimator

- Censored individuals have the same prospect of survival as those who continue to be followed. This can not be tested for and can lead to a bias that artificially reduces S.
- Survival prospects are the same for early as for late recruits to the study (can be tested for).
- The event studied (e.g. death) happens at the specified time. Late recording of the event studied will cause artificial inflation of S

The Kaplan-Meier Estimator

- Aim: estimate the Survival Function $S(t)$ based on a sample of failure times $T_1, \ldots, T_n$
  - ▷ Remember: $S(t)$ is the probability of being alive at time t

- If there was no censoring, we could simply

$$\hat{S}(t) = \frac{\text{number of patients alive at time } t}{n} = \frac{1}{n}\sum_{i=1}^{n} I(T_i > t)$$

where $I(T_i > t)$ equals 1 if $T_i > t$, and 0 otherwise

- However, we do have censored observations

The Kaplan-Meier Estimator

- To take into account censoring in the estimation of $S(t)$ we will use *the law of total probability*

- For instance, the probability of surviving 2 years can be computed as:

$$\begin{aligned} S(2) &= \Pr(T_i > 2) \\ &= \Pr(T_i > 1) \times \Pr(T_i > 2 \mid T_i > 1) \end{aligned}$$

- In words, the probability of surviving year 2 is the product of
  - ▷ the probability of surviving year 1 and
  - ▷ the conditional probability of surviving up to year 2 given still being alive at year 1

The Kaplan-Meier Estimator

- So $S(2)$ can be estimated by

$$\hat{S}(2) = \frac{\#\text{ patients alive at year 1}}{\#\text{ patients at risk up to year 1}} \times \frac{\#\text{ patients alive at year 2}}{\#\text{ patients at risk up to year 2}}$$

- If we apply this idea repeatedly, we can obtain survival probabilities for every time point t

The Kaplan-Meier Estimator

- Let $t_1, t_2, \ldots, t_k$ denote the unique event times in the sample at hand

- We account for censoring by suitably adjusting the risk set $\Rightarrow$ the *Kaplan-Meier Estimator*

$$\hat{S}_{KM}(t) = \prod_{i: t_i \leq t} \frac{r_i - d_i}{r_i}$$

where d_i is the number of events at time t_i, and r_i the number of patients still at risk at time t_i

- ▷ still at risk means alive and not censored

The Kaplan-Meier Estimator

- First remove from the list of unordered survival times all those times that are censored; we are thus working only with those times at which people failed.
- Then order the remaining failure times from smallest to largest, and count ties only once.
- The third column gives frequency counts, denoted by r_i, of those persons who are at risk in the time starting with failure time $t_{(i)}$.
- The fourth column gives frequency counts, denoted by d_i, of those persons who are at event occurred for them before the time $t_{(i)}$.

- A small example

$$1 \;\; 5^+ \;\; 6 \;\; 6 \;\; 8 \;\; 8^+ \;\; 9 \;\; 11^+$$

$+$ denotes a censored time

i	t_i	r_i	d_i	$(r_i - d_i)/r_i$
1	1	8	1	7/8
2	6	6	2	4/6
3	8	4	1	3/4
4	9	2	1	1/2

Note: We have not counted person #5 (who was censored at week 8) here because this person's censored time is exactly at the end of the interval. We count this person in the following interval.

The Kaplan-Meier Estimator

The K-M estimated percent survival of the subject(s) at each point in time $t_{(i)}$

$$
\begin{aligned}
\hat{S}_{KM}(t) &= 1, \quad 0 \le t < 1 \\
&= 7/8 = 0.875, \quad 1 \le t < 6 \\
&= (7/8)(4/6) = 0.583, \quad 6 \le t < 8 \\
&= (7/8)(4/6)(3/4) = 0.438, \quad 8 \le t < 9 \\
&= (7/8)(4/6)(3/4)(1/2) = 0.219, \quad 9 \le t < 11
\end{aligned}
$$

Note: the estimate of $S(t)$ is undefined for $t > 11$ since not all subjects have died by $t = 11$

The Kaplan-Meier Estimator

- The variance of $\hat{S}_{KM}(t)$ can be estimated using Greenwood's formula

- Using the formula and asymptotic normality of $\hat{S}_{KM}(t)$, we can derive a 95% confidence interval

- Problem: this can exceed 1 or fall below 0!

- A better asymmetric 95% confidence interval for $\hat{S}_{KM}(t)$ that respects the boundaries is derived from a symmetric 95% confidence interval for either

$$\hat{H}_{KM}(t) = -\log \hat{S}_{KM}(t) \quad \text{or} \quad \log \hat{H}_{KM}(t) = \log\{-\log \hat{S}_{KM}(t)\}$$

The Kaplan-Meier Estimator

Greenwood's formula (Collett 2.1.3)

We can think of the KM estimator as

$$\hat{S}(t) = \prod_{j:\tau_j<t} (1 - \hat{\lambda}_j)$$

where $\hat{\lambda}_j = d_j/r_j$.

Since the $\hat{\lambda}_j$'s are just binomial proportions, we can apply standard likelihood theory to show that each $\hat{\lambda}_j$ is approximately normal, with mean the true λ_j, and

$$var(\hat{\lambda}_j) \approx \frac{\hat{\lambda}_j(1 - \hat{\lambda}_j)}{r_j}$$

Also, the $\hat{\lambda}_j$'s are independent in large enough samples.

Since $\hat{S}(t)$ is a function of the λ_j's, we can estimate its variance using the **delta method**:

Delta method: If Y is normal with mean μ and variance σ^2, then $g(Y)$ is approximately normally distributed with mean $g(\mu)$ and variance $[g'(\mu)]^2\sigma^2$.

Two specific examples of the delta method:

(A) $Z = log(Y)$

$$\text{then} \quad Z \sim N\left[log(\mu), \left(\frac{1}{\mu}\right)^2 \sigma^2\right]$$

(B) $Z = exp(Y)$

$$\text{then} \quad Z \sim N\left[e^{\mu}, [e^{\mu}]^2\sigma^2\right]$$

The Kaplan-Meier Estimator

Greenwood's formula (continued)

Instead of dealing with $\hat{S}(t)$ directly, we will look at its log:

$$\log[\hat{S}(t)] = \sum_{j:\tau_j<t} \log(1-\hat{\lambda}_j) \qquad \text{where } \hat{\lambda}_j = d_j/r_j.$$

Thus, by approximate independence of the $\hat{\lambda}_j$'s,

$$\begin{aligned} var(\log[\hat{S}(t)]) &= \sum_{j:\tau_j<t} var[\log(1-\hat{\lambda}_j)] \\ \text{by (A)} \quad &= \sum_{j:\tau_j<t} \left(\frac{1}{1-\hat{\lambda}_j}\right)^2 var(\hat{\lambda}_j) \\ &= \sum_{j:\tau_j<t} \left(\frac{1}{1-\hat{\lambda}_j}\right)^2 \hat{\lambda}_j(1-\hat{\lambda}_j)/r_j \\ &= \sum_{j:\tau_j<t} \frac{\hat{\lambda}_j}{(1-\hat{\lambda}_j)r_j} \\ &= \sum_{j:\tau_j<t} \frac{d_j}{(r_j-d_j)r_j} \end{aligned}$$

Now, $\hat{S}(t) = \exp[\log[\hat{S}(t)]]$. Thus by (B),

$$var(\hat{S}(t)) = [\hat{S}(t)]^2 var\left[\log[\hat{S}(t)]\right]$$

Greenwood's Formula:

$$var(\hat{S}(t)) = [\hat{S}(t)]^2 \Sigma_{j:\tau_j<t} \frac{d_j}{(r_j-d_j)r_j}$$

Dr Rah

Time	Risk(r)	Fail(d)	Lost	Fun	K-M	Error	Conf.L	Conf.U
1	8	1	0	7/8	0.875	0.117	0.646	**1.104**
6	6	2	1	4/6	0.5833	0.186	0.219	0.948
8	4	1		3/4	0.4375	0.188	0.069	0.806
9	2	1	1	1/2	0.21875	0.181	**-0.136**	0.574

$1\quad 5^+\quad 6\quad 6\quad 8\quad 8^+\quad 9\quad 11^+$

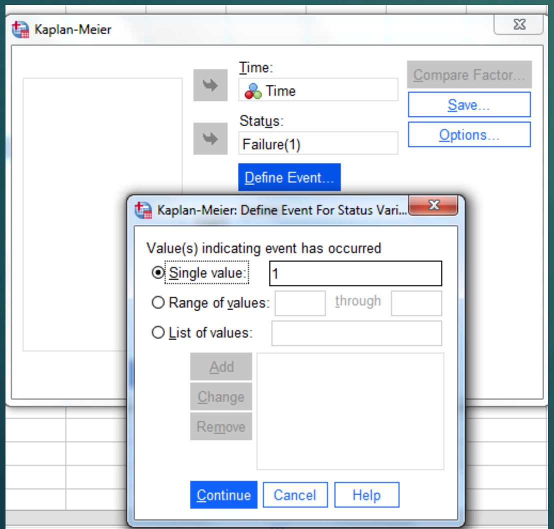
Greenwood's Formula

- $V[S(T)] = S(T_t)^2 \sum_{j=1}^{t-1} \frac{d_j}{(r_j - d_j) r_j}$
- $V[S(T_1)] = S(T_1)^2 \times \frac{d_1}{(r_1 - r_1) r_1} = 0.875^2 \left(\frac{1}{7\times 8}\right) = 0.01367$ **error=0.117**
- $V[S(T_2)] = S(T_2)^2 \left(\frac{d_2}{(r_2 - d_2) r_2} + \frac{d_1}{(r_1 - d_1) r_1}\right) = 0.5833^2 \times \left(\frac{2}{6\times 4} + \frac{1}{7\times 8}\right) = 0.03443$ **error=0.186**
- $V[S(T_3)] = S(T_3)^2 \left(\frac{d_3}{(r_3 - d_3) r_3} + \frac{d_2}{(r_2 - d_2) r_2} + \frac{d_1}{(r_1 - d_1) r_1}\right) = 0.4375^2 \times \left(\frac{1}{3\times 4} + \frac{2}{6\times 4} + \frac{1}{7\times 8}\right)$
$= 0.03532$ **error=0.188**
- $V[S(T_4)] = S(T_4)^2 \left(\frac{d_4}{(r_4 - d_4) r_4} + \frac{d_3}{(r_3 - d_3) r_3} + \frac{d_2}{(r_2 - d_2) r_2} + \frac{d_1}{(r_1 - d_1) r_1}\right) = 0.21875^2 \left(\frac{1}{1\times 2} + \frac{1}{3\times 4} + \frac{2}{6\times 4}\right.$
$\left. + \frac{1}{7\times 8}\right) = 0.032756$ **error=0.181**

SPSS

	Time	Failure
1	1.00	1.00
2	5.00	.00
3	6.00	1.00
4	6.00	1.00
5	8.00	1.00
6	8.00	.00
7	9.00	1.00
8	11.00	.00

- ورود داده
- تعریف متغیر شکست از مسیر
- Analyze/ Survival/ Kaplan-Meier
- پس از کلیک روی گزینه Continue روی گزینه option کلیک
- کرده و گزینه Survival Table را تیک می زنیم.
- در خروجی تابع بقا و انحراف معیار تابع بقا به روش گرین وود
- گزارش خواهد.



Dr Rahimzadeh, Biostatistics

Out Put

Survival Table

	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	1.000	1.00	.875	.117	1	7
2	5.000	.00	.	.	1	6
3	6.000	1.00	.	.	2	5
4	6.000	1.00	.583	.186	3	4
5	8.000	1.00	.438	.188	4	3
6	8.000	.00	.	.	4	2
7	9.000	1.00	.219	.181	5	1
8	11.000	.00	.	.	5	0

SPSS

- در صورتی که بخواهید تابع بقا کاپلان-مایر و انحراف معیار آن در فایل داده ذخیره شود در گزینه save کنار دو گزینه Survival و Standard Error of Survival تیک می زنید.

	Time	Failure	SUR_1	SE_1
1	1.00	1.00	.87500	.11693
2	5.00	.00	.	.
3	6.00	1.00	.58333	.18556
4	6.00	1.00	.58333	.18556
5	8.00	1.00	.43750	.18793
6	8.00	.00	.	.
7	9.00	1.00	.21875	.18098
8	11.00	.00	.	.

The Kaplan-Meier Estimator with Confidence Interval

The standard error reported is given by Greenwood's formula (Greenwood 1926):

$$\widehat{\mathrm{Var}}\{\widehat{S}(t)\} = \widehat{S}^{\,2}(t) \sum_{j|t_j \le t} \frac{d_j}{n_j(n_j - d_j)}$$

(Kalbfleisch and Prentice 2002, 17–18). These standard errors, however, are not used for confidence intervals. Instead, the asymptotic variance of $\ln[-\ln \widehat{S}(t)]$,

$$\widehat{\sigma}^{\,2}(t) = \frac{\sum \frac{d_j}{n_j(n_j-d_j)}}{\{\sum \ln(\frac{n_j-d_j}{n_j})\}^2}$$

is used, where sums are calculated over $j|t_j \le t$ (Kalbfleisch and Prentice 2002, 18). The confidence bounds are then $\widehat{S}(t)^{\exp(\pm z_{\alpha/2}\widehat{\sigma}(t))}$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the normal distribution. `sts` suppresses reporting the standard error and confidence bounds if the data are `pweighted` because these formulas are no longer appropriate.

Note: SPSS software does not provide confidence intervals

Time	Risk	Fail	Lost	K-M	-LOG(K-M)	Error(-LOG(K_M))	Conf.L	Conf.U
1	8	1	0	0.875	0.0580	1.00074	0.387	0.9814
6	6	2	1	0.5833	0.2341	0.5902	0.1801	0.8441
8	4	1		0.4375	0.3590	0.5196	0.1014	0.7149
9	2	1	1	0.21875	0.6601	0.5444	0.0121	0.5928

$$V[\ln(-\ln(S(t_i))] = \frac{\sum_{j=1}^{t-1}\frac{d_j}{(r_j-d_j)r_j}}{\left[\sum_{j=1}^{t-1} ln(\frac{r_j-d_j}{r_j})\right]^2}$$

$(\widehat{S(t_1)})^{\exp(\pm Z\alpha/_2\widehat{\sigma(t)})}$ Kalbfleisch Formula

$V[(\ln(-\ln(\widehat{S(t_1)}))] = \frac{\frac{1}{7\times 8}}{[Ln(\frac{7}{8})]^2}$=1.00149 Error=1.00074 $0.875^{\exp(1.96*1.00074)}$=**0.387** , $0.875^{\exp(-1.96*1.00074)}$=**0.9814**

$V[(\ln(-\ln(\widehat{S(t_2)}))] = \frac{\frac{2}{4\times 6}+\frac{1}{7\times 8}}{[\ln(\frac{4}{6})+Ln(\frac{7}{8})]^2}$=0.3483, Error=0.5902 ($0.5833^{\exp(1.96*0.5902)} = \mathbf{0.1801}$, $0.5833^{\exp(-1.96*0.5902)} = \mathbf{0.8441}$)

$V[(\ln(-\ln(\widehat{S(t_3)}))] = \frac{\frac{1}{3\times 4}+\frac{2}{4\times 6}+\frac{1}{7\times 8}}{[\ln(\frac{3}{4})+\ln(\frac{4}{6})+Ln(\frac{7}{8})]^2}$=0.27 Error=0.5196, ($0.4375^{\exp(1.96*0.5196)} = \mathbf{0.1014}$, $0.4375^{\exp(-1.96*0.5196)}$=**0.7149**

$V[(\ln(-\ln(\widehat{S(t_4)}))] = \frac{\frac{1}{1\times 2}+\frac{1}{3\times 4}+\frac{2}{4\times 6}+\frac{1}{7\times 8}}{[\ln(\frac{1}{2})+\ln(\frac{3}{4})+\ln(\frac{4}{6})+Ln(\frac{7}{8})]^2}$=0.2963 E=0.5444,

$0.21875^{\exp(1.96*0.5444)} = \mathbf{0.0121}$, $0.21875^{\exp(-1.96*0.5444)} = \mathbf{0.5928}$

Example of calculation Life Table

- McIllmurray and Turkie (2) describe a clinical trial of 25 patients for the treatment and 24 patients for the Control of Dukes' C colorectal cancer. The data for the two treatments, linoleic acid or control are given.

Table 12.1 Survival in 49 patients with Dukes' C colorectal cancer randomly assigned to either γ linoleic acid or control treatment

Treatment	Survival time (months)
γ linoleic acid (n=25)	1+, 5+, 6, 6, 9+, 10, 10, 10+, 12, 12, 12, 12, 12+, 13+, 15+, 16+, 20+, 24, 24+, 27+, 32, 34+, 36+, 36+, 44+
Control (n=24)	3+, 6, 6, 6, 6, 8, 8, 12, 12, 12+, 15+, 16+, 18+, 18+, 20, 22+, 24, 28+, 28+, 28+, 30, 30+, 33+, 42

- The + sign indicates censored data.
- We calculate K-M and Confidence Interval for the <u>treatment group</u>.

KM with CI

Time	At risk	Fail	survival	error	[95% conf. int.]	
6	23	2	0.9130	0.0588	0.6949	0.9775
10	20	2	0.8217	0.0809	0.5917	0.9292
12	17	4	0.6284	0.1048	0.3911	0.7946
24	8	1	0.5498	0.1175	0.2998	0.7431
32	5	1	0.4399	0.1360	0.1794	0.6753

- $\hat{S}_{KM}(t_i) = \prod_{j|t_j\le t}\left(\frac{r_j - d_j}{d_j}\right)$
- $\hat{S}_{KM}(T_1 = 6) = \frac{21}{23}$=0.9130
- $\hat{S}_{KM}(T_2 = 10) = \frac{21}{23} \times \frac{18}{20} = \mathbf{0.8217}$
- $\hat{S}_{KM}(T_3 = 12) = \frac{21}{23} \times \frac{18}{20} \times \frac{13}{17} = \mathbf{0.6284}$
- $\hat{S}_{KM}(T_4 = 24) = \frac{21}{23} \times \frac{18}{20} \times \frac{13}{17} \times \frac{7}{8} = \mathbf{0.5498}$
- $\hat{S}_{KM}(T_5 = 32) = \frac{21}{23} \times \frac{18}{20} \times \frac{13}{17} \times \frac{7}{8} \times \frac{4}{5} = \mathbf{0.4399}$

Error for KM

Time	At Risk	Fail	function	error	[95% conf. int.]	
6	23	2	0.9130	0.0588	0.6949	0.9775
10	20	2	0.8217	0.0809	0.5917	0.9292
12	17	4	0.6284	0.1048	0.3911	0.7946
24	8	1	0.5498	0.1175	0.2998	0.7431
32	5	1	0.4399	0.1360	0.1794	0.6753

$$V[S(T)] = S(T_t)^2 \sum_{j=1}^{t-1} \frac{d_j}{(r_j - d_j) r_j}$$

$$V[S(T_1)] = S(T_1)^2 \times \frac{d_1}{(n_1 - r_1) r_1} = 0.9130^2 \left(\frac{2}{21 \times 23}\right) = 0.00345$$ **error=0.0588**

$$V[S(T_2)] = S(T_2)^2 \left(\frac{d_2}{(n_2 - d_2) r_2} + \frac{d_1}{(n_1 - d_1) r_1}\right) = 0.8217^2 \times \left(\frac{2}{18 \times 20} + \frac{2}{21 \times 23}\right) = 0.00655$$

error=0.0809

$$V[S(T_3)] = S(T_3)^2 \left(\frac{d_3}{(n_3 - d_3) r_3} + \frac{d_2}{(n_2 - d_2) r_2} + \frac{d_1}{(n_1 - d_1) r_1}\right) = 0.6284^2 \times \left(\frac{2}{13 \times 17} + \frac{2}{18 \times 20} + \frac{2}{21 \times 23}\right) = 0.01098$$

error=0. 1048

Error for KM

Time	At Risk	Fail	function	error	[95% conf. int.]	
6	23	2	0.9130	0.0588	0.6949	0.9775
10	20	2	0.8217	0.0809	0.5917	0.9292
12	17	4	0.6284	0.1048	0.3911	0.7946
24	8	1	0.5498	0.1175	0.2998	0.7431
32	5	1	0.4399	0.1360	0.1794	0.6753

- $V[S(T_4)] = S(T_4)^2\left(\frac{d_4}{(n_4-d_4)r_4}+\frac{d_3}{(n_3-d_3)r_3}+\frac{d_2}{(n_2-d_2)r_2}+\frac{d_1}{(n_1-d_1)r_1}\right) = 0.5498^2\left(\frac{1}{7\times8}+\frac{2}{13\times17}+\frac{2}{18\times20}+\frac{2}{21\times23}\right) = 0.0138$ **error=0.1175**
- $V[S(T_5)] = S(T_5)^2\left(\frac{d_5}{(n_5-d_5)r_5}+\frac{d_4}{(n_4-d_4)r_4}+\frac{d_3}{(n_3-d_3)r_3}+\frac{d_2}{(n_2-d_2)r_2}+\frac{d_1}{(n_1-d_1)r_1}\right) = 0.4399^2\left(\frac{1}{4\times5}+\frac{2}{13\times17}+\frac{2}{18\times20}+\frac{2}{21\times23}+\frac{1}{7\times8}+\frac{2}{13\times17}+\frac{2}{18\times20}+\frac{2}{21\times23}\right)=0.0185$ **error=0.1360**

STATA

Dr Rahimzadeh, Biostatistics

STATA

- Data entry is quite similar to SPSS.
- You can open data files of various formats from the import section in STATA.
- Survival analysis is one of the important menus in the analysis section.
- First of all, you need to tell the Stata that the data file is a survival data file.
- To do this, we use the sts command as follows.
- stset time, failure(failure==1)

STATA: sts list

Input Data

- A small example

$$1 \;\; 5^+ \;\; 6 \;\; 6 \;\; 8 \;\; 8^+ \;\; 9 \;\; 11^+$$

	time	failure	_st	_d	_t	_t0
1	1	1	1	1	1	0
2	5	0	1	0	5	0
3	6	1	1	1	6	0
4	6	1	1	1	6	0
5	8	1	1	1	8	0
6	8	0	1	0	8	0
7	9	1	1	1	9	0
8	11	0	1	0	11	0

KM Out Put

Kaplan-Meier survivor function

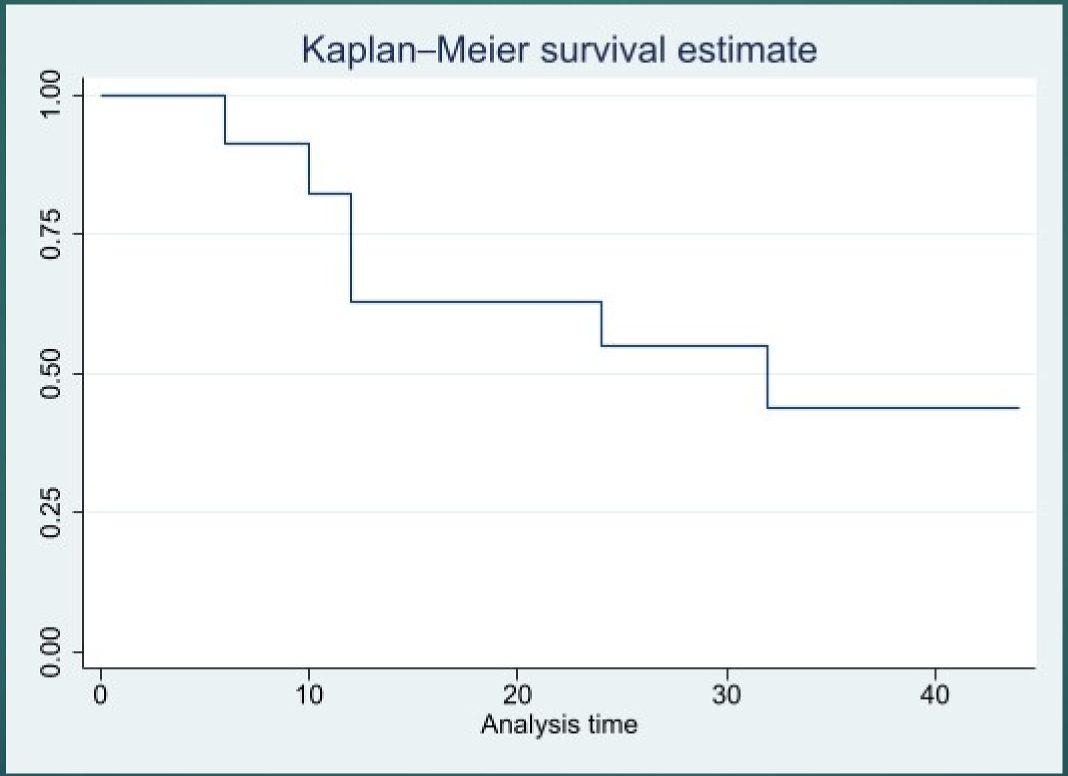
Time	At risk	Fail	Lost	Survivor function	Std. error	[95% conf.	int.]
1	8	1	0	0.8750	0.1169	0.3870	0.9814
5	7	0	1	0.8750	0.1169	0.3870	0.9814
6	6	2	0	0.5833	0.1856	0.1802	0.8441
8	4	1	1	0.4375	0.1879	0.1014	0.7419
9	2	1	0	0.2188	0.1810	0.0121	0.5928
11	1	0	1	0.2188	0.1810	0.0121	0.5928

STATA Out PUT: McIllmurray and Turkie, treatment group

Kaplan-Meier survivor function

Time	At risk	Fail	Lost	Survivor function	Std. error	[95% conf.	int.]
1	25	0	1	1.0000	.	.	.
5	24	0	1	1.0000	.	.	.
6	23	2	0	0.9130	0.0588	0.6949	0.9775
9	21	0	1	0.9130	0.0588	0.6949	0.9775
10	20	2	1	0.8217	0.0809	0.5917	0.9292
12	17	4	1	0.6284	0.1048	0.3911	0.7946
13	12	0	1	0.6284	0.1048	0.3911	0.7946
15	11	0	1	0.6284	0.1048	0.3911	0.7946
16	10	0	1	0.6284	0.1048	0.3911	0.7946
20	9	0	1	0.6284	0.1048	0.3911	0.7946
24	8	1	1	0.5498	0.1175	0.2998	0.7431
27	6	0	1	0.5498	0.1175	0.2998	0.7431
32	5	1	0	0.4399	0.1360	0.1794	0.6753
34	4	0	1	0.4399	0.1360	0.1794	0.6753
36	3	0	2	0.4399	0.1360	0.1794	0.6753
44	1	0	1	0.4399	0.1360	0.1794	0.6753

STATA: sts graph

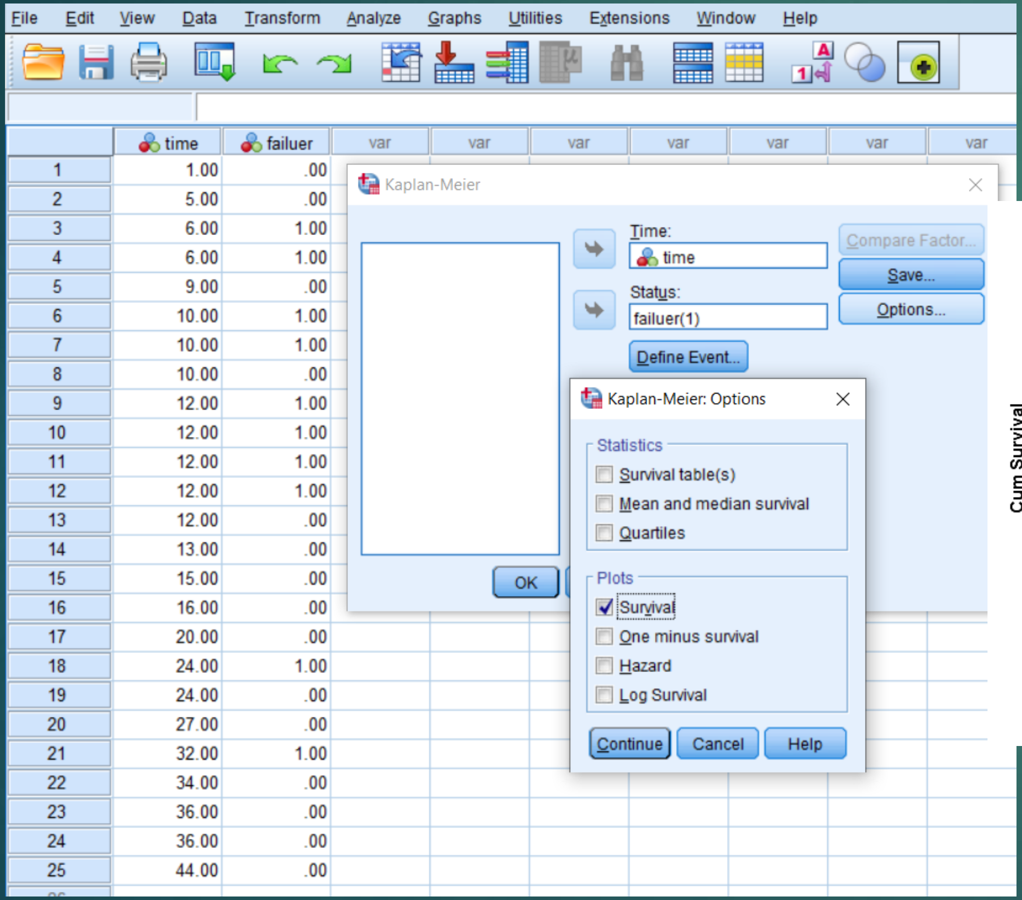


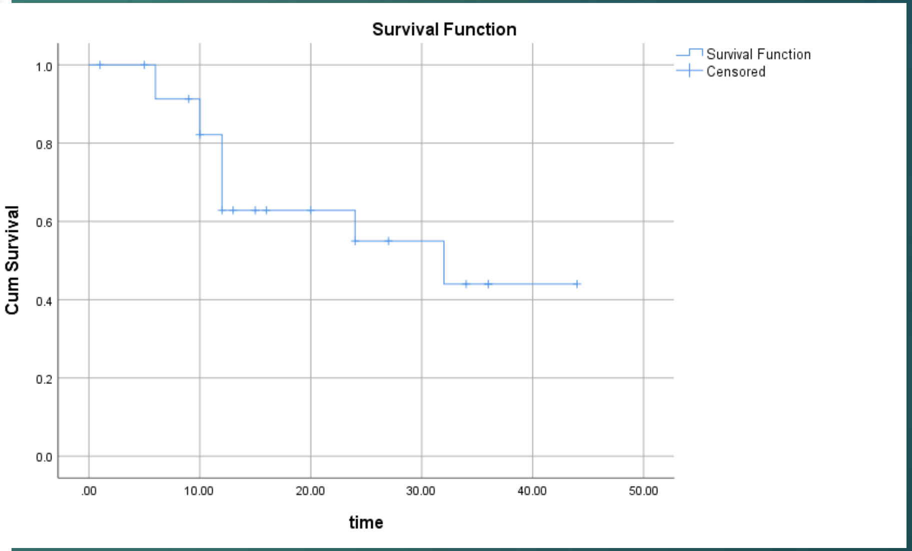
SPSS: McIllmurray and Turkie, treatment group

time	failure	SUR_1	SE_1
6.00	1.00	.91304	.05875
10.00	1.00	.82174	.08092
12.00	1.00	.62839	.10477
24.00	1.00	.54984	.11748
32.00	1.00	.43987	.13604

Note: SPSS software does not provide confidence intervals.

SPSS: K-M





Dr Rahimzadeh, Biostatistics

Practice

- Calculate the Confidence Interval for Survival Function for previous example with the treatment group with the asymptotic of variance.

Disadvantages of Kaplan-Meier Method

- In most statistical Package for handling a censoring that happens at the same time as a failure is to assume that the failure occurred before the censoring. Still, in this situation overestimation of Product Limit estimator function in the presence of ties.
- There is a level at which delayed entries cause considerable problems. In these entries' presence, the Kaplan-Meier procedure for calculating the survivor curve can yield absurd results. This happens when some late arrivals enter the study after everyone before them has failed.
- Few datasets actually have such gaps. Gaps in subject histories are not a problem; it is only when all the gaps in all the subjects add up in such a way that no one is at risk during a period that the dataset itself will have such a gap.
- When datasets do have such gaps, you want to avoid the Kaplan-Meier estimate and use the Nelson-Aalen, and still, you need to think carefully as you interpret results.

Limitations of Kaplan-Meier

- Mainly descriptive
- Doesn’t control for covariates
- Requires categorical predictors
- Can’t accommodate time-dependent variables