

Biost 517 / Biost 514

Applied Biostatistics I / Biostatistics I



Discussion Week 10:
Nonparametric Survival Analysis

Outline



- Kaplan-Meier estimators
 - Fitting K-M curves in R
 - Estimating survival probabilities and quantiles
 - Comparing Survival Functions with Log-rank Test

Kaplan-Meier plots in R



The Kaplan-Meier estimator is a non-parametric estimator of the survival function from survival data.

We can use the Kaplan-Meier estimate of survival functions to:

- Estimate survival probabilities for a given time
- Estimate survival time for a given quantile
- Easily compare survival curves for different (categorical) groups

Data



We will be using the PSA data set available on Canvas.

Variables of interest:

- `inrem`: indicator of remission status
- `obstime`: observation time (until relapse or end of follow-up)
- `grade`: tumor grade (1=least aggressive, 3=most aggressive)

This week, we will be

- summarizing relapse probabilities and time until relapse
- examining the relationship between tumor grade and relapse

Kaplan-Meier plots in R: setting up



We start by loading in the 'survival' package and creating the event indicator:

```
## Load the required package, "survival"
library("survival")

# The event indicator is 'inrem': 1 = no, 2 = yes.
# Recode event as 1 if patient is not in remission (relapse) at observed time.
psa$event <- ifelse(psa$inrem == "no", 1, 0)
```

- event = 1 if the event (discontinuation of remission) is observed, event = 0 if the event is not observed (still in remission).
- End of remission (event = 1) is equivalent to when 'inrem' is 'no'

K-M plots: Survival object



`Surv()` creates the survival object from the observed times and event indicator; this will be our outcome variable in our Kaplan-Meier estimator and proportional hazards model.

```
> survobj <- with(psa, Surv(obstime, event))
```

```
> survobj
```

```
[1] 42+ 48+ 40+ 75+ 30+ 24+ 58+ 36+ 40  60+ 60  48  30  60+ 48+ 35+ 42+  
[18] 45  43  42  40  39  36  26  26  22  21  20  18  17  16  16  15  12  
[35]  8   9   7   6   3   3   3   6   6   1  31  10  14  12  32  42+
```

Q. What does “+” indicate about the observations?

K-M plots: Fitting the K-M curve



`survfit()` computes the Kaplan-Meier estimate from the survival object created earlier.

- The 1 in the formula (`survobj ~ 1`) indicates we are estimating the survival function for one group
- More on the summary later...

```
> eS <- survfit( survobj ~ 1 )
```

```
> summary(eS)
```

```
Call: survfit(formula = survobj ~ 1)
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	50	1	0.980	0.0198	0.9420	1.000
3	49	3	0.920	0.0384	0.8478	0.998
6	46	3	0.860	0.0491	0.7690	0.962
7	43	1	0.840	0.0518	0.7443	0.948
8	42	1	0.820	0.0543	0.7201	0.934
9	41	1	0.800	0.0566	0.6965	0.919
10	40	1	0.780	0.0586	0.6732	0.904

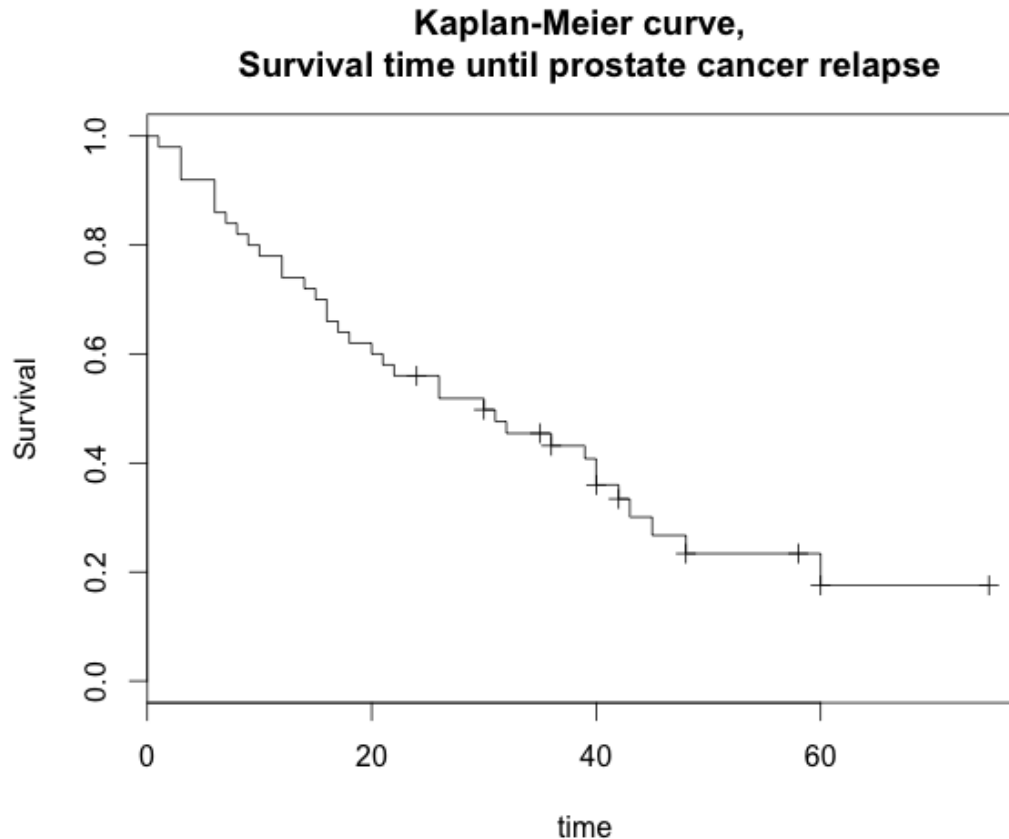
K-M plots: Plotting the K-M curve



Use `plot()` on the estimated survival function from `survfit()` :

```
# plot the resulting KM curve
```

```
> plot(eS, main="Kaplan-Meier curve, \n Survival time until cancer relapse",  
      xlab="time", ylab="Survival", conf.int=FALSE, mark.time = TRUE)
```

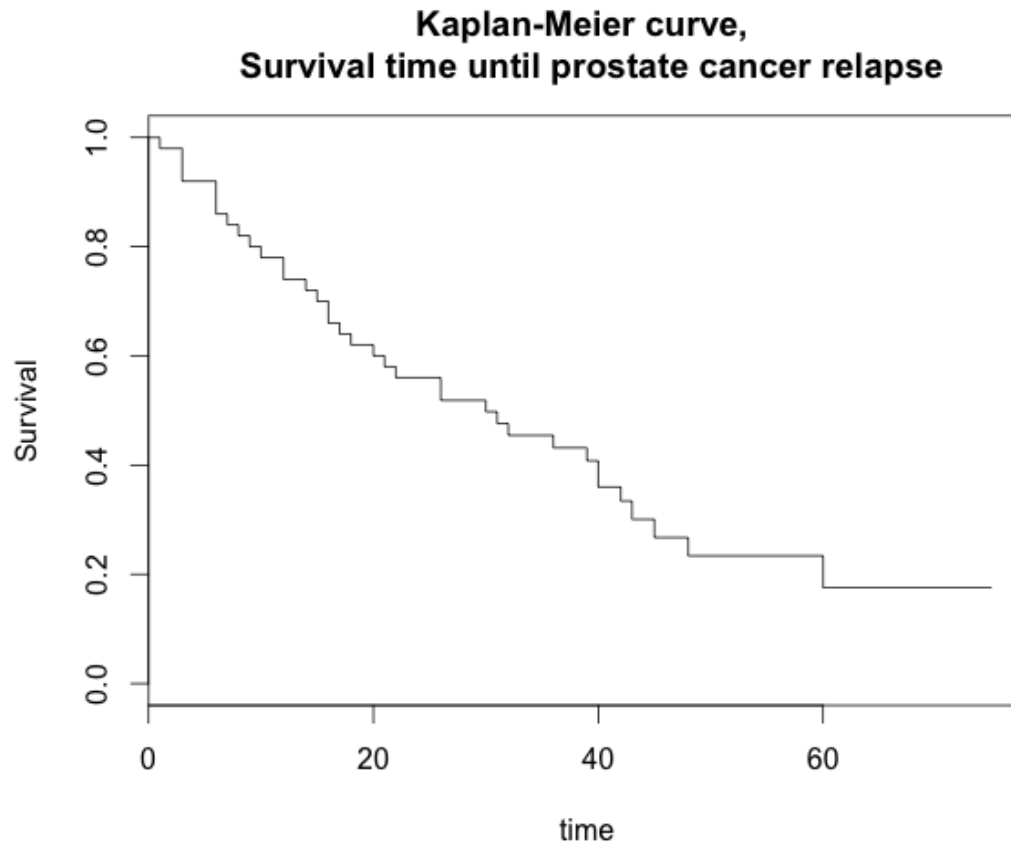


K-M plots: Plotting the K-M curve



To exclude censoring times, set `mark.times=FALSE`:

```
> plot(eS, main="Kaplan-Meier curve, \n Survival time until cancer relapse",  
      xlab="time", ylab="Survival", conf.int=FALSE, mark.time = FALSE)
```

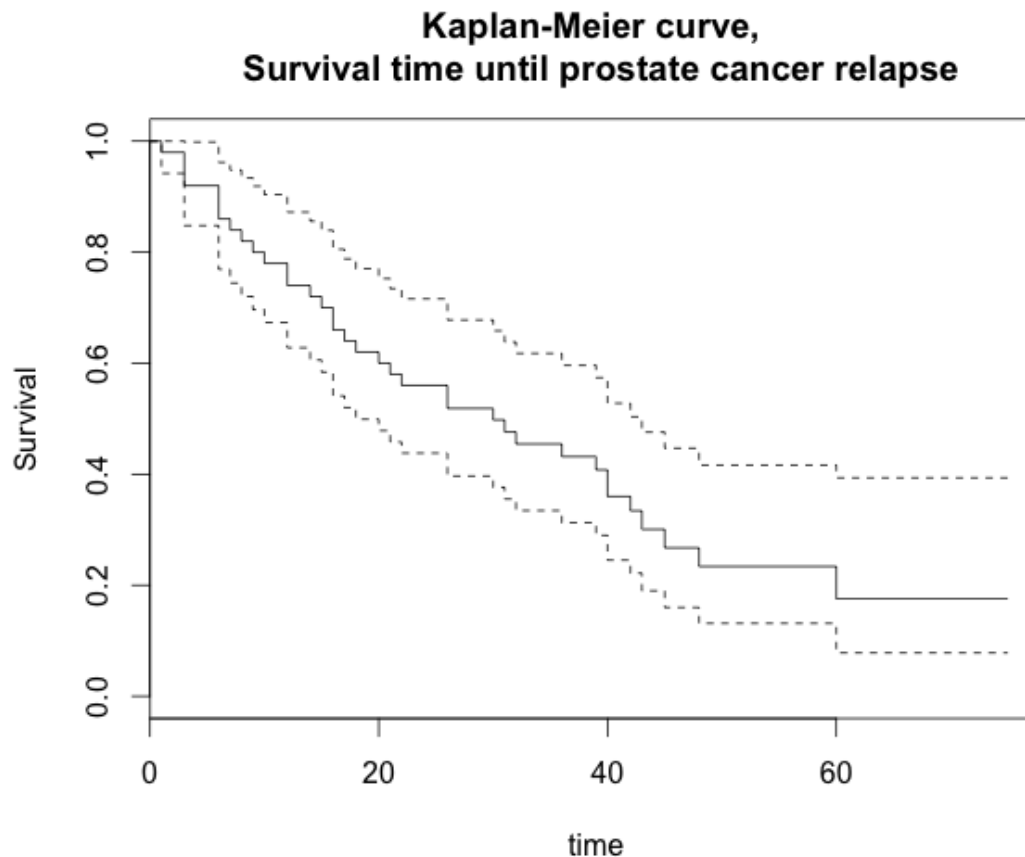


K-M plots: Plotting the K-M curve



To include a 95% CI, set `conf.int=TRUE`:

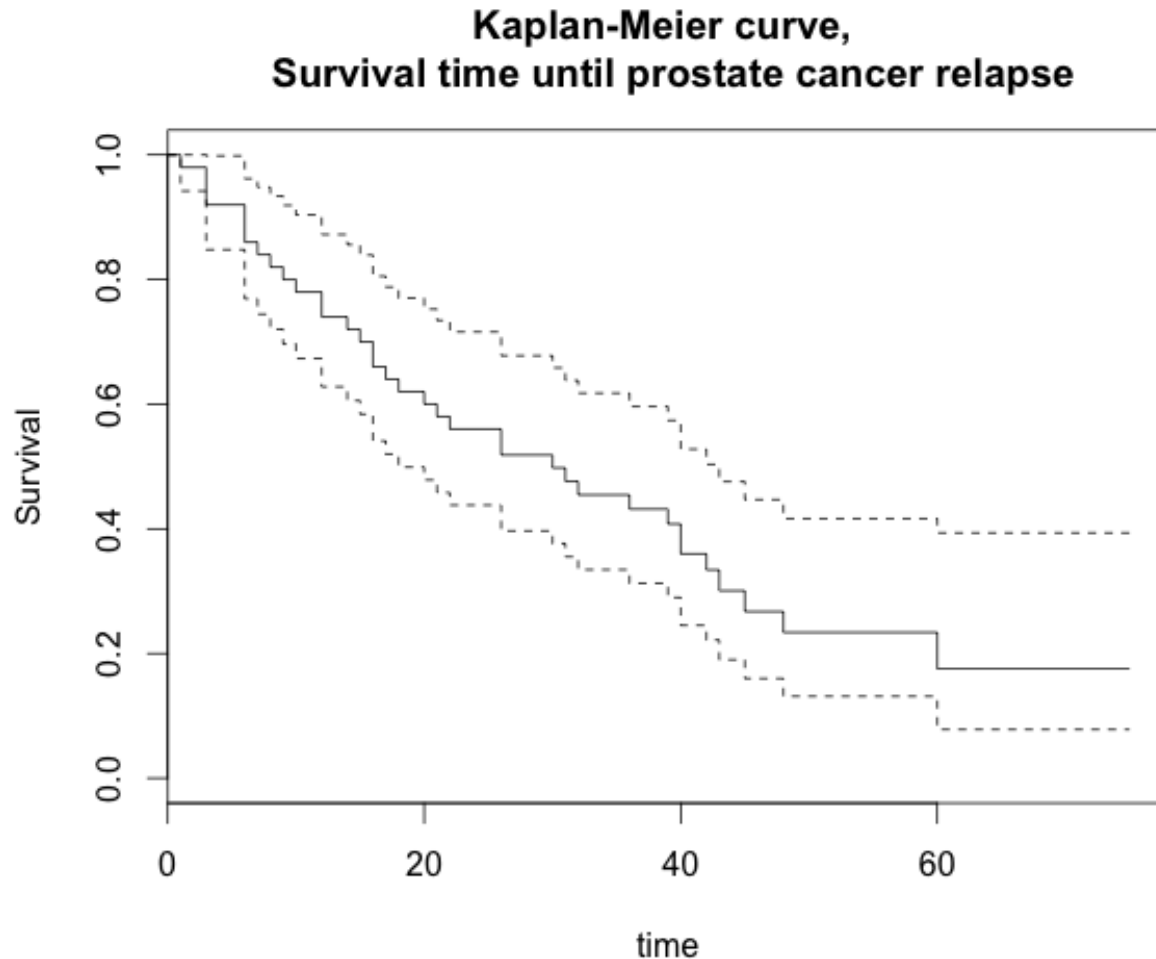
```
> plot(eS, main="Kaplan-Meier curve, \n Survival time until cancer relapse",  
      xlab="time", ylab="Survival", conf.int=TRUE, mark.time = FALSE)
```



Estimating survival probabilities



Q. What is the estimated probability of remaining in remission at two years?

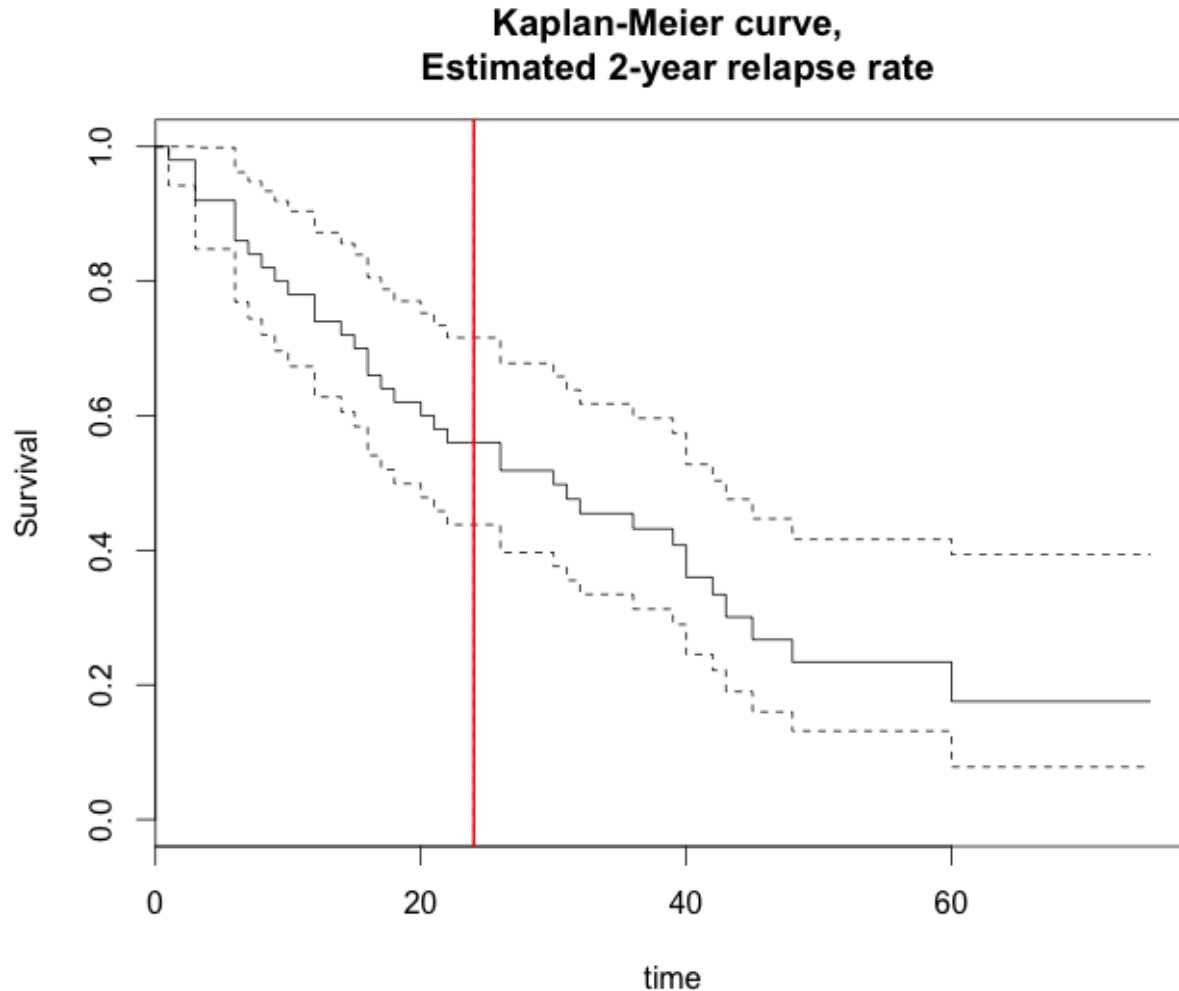


Estimating survival probabilities



Q. What is the estimated probability of remaining in remission at two years?

Visually:



Estimating survival probabilities



Q. What is the estimated probability of remaining in remission at two years?

```
> summary(eS)
```

```
Call: survfit(formula = survobj ~ 1)
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	50	1	0.980	0.0198	0.9420	1.000
3	49	3	0.920	0.0384	0.8478	0.998
6	46	3	0.860	0.0491	0.7690	0.962
7	43	1	0.840	0.0518	0.7443	0.948
8	42	1	0.820	0.0543	0.7201	0.934
9	41	1	0.800	0.0566	0.6965	0.919
10	40	1	0.780	0.0586	0.6732	0.904
12	39	2	0.740	0.0620	0.6279	0.872
14	37	1	0.720	0.0635	0.6057	0.856
15	36	1	0.700	0.0648	0.5838	0.839
16	35	2	0.660	0.0670	0.5409	0.805
17	33	1	0.640	0.0679	0.5199	0.788
18	32	1	0.620	0.0686	0.4991	0.770
20	31	1	0.600	0.0693	0.4785	0.752
21	30	1	0.580	0.0698	0.4581	0.734
22	29	1	0.560	0.0702	0.4380	0.716
26	27	2	0.519	0.0709	0.3967	0.678
30	25	1	0.498	0.0710	0.3764	0.658

Estimating survival probabilities



Q. What is the estimated probability of no remission at two years?
From the summary output: **.560; 95% CI: (.438, .716)**

```
> summary(eS)
```

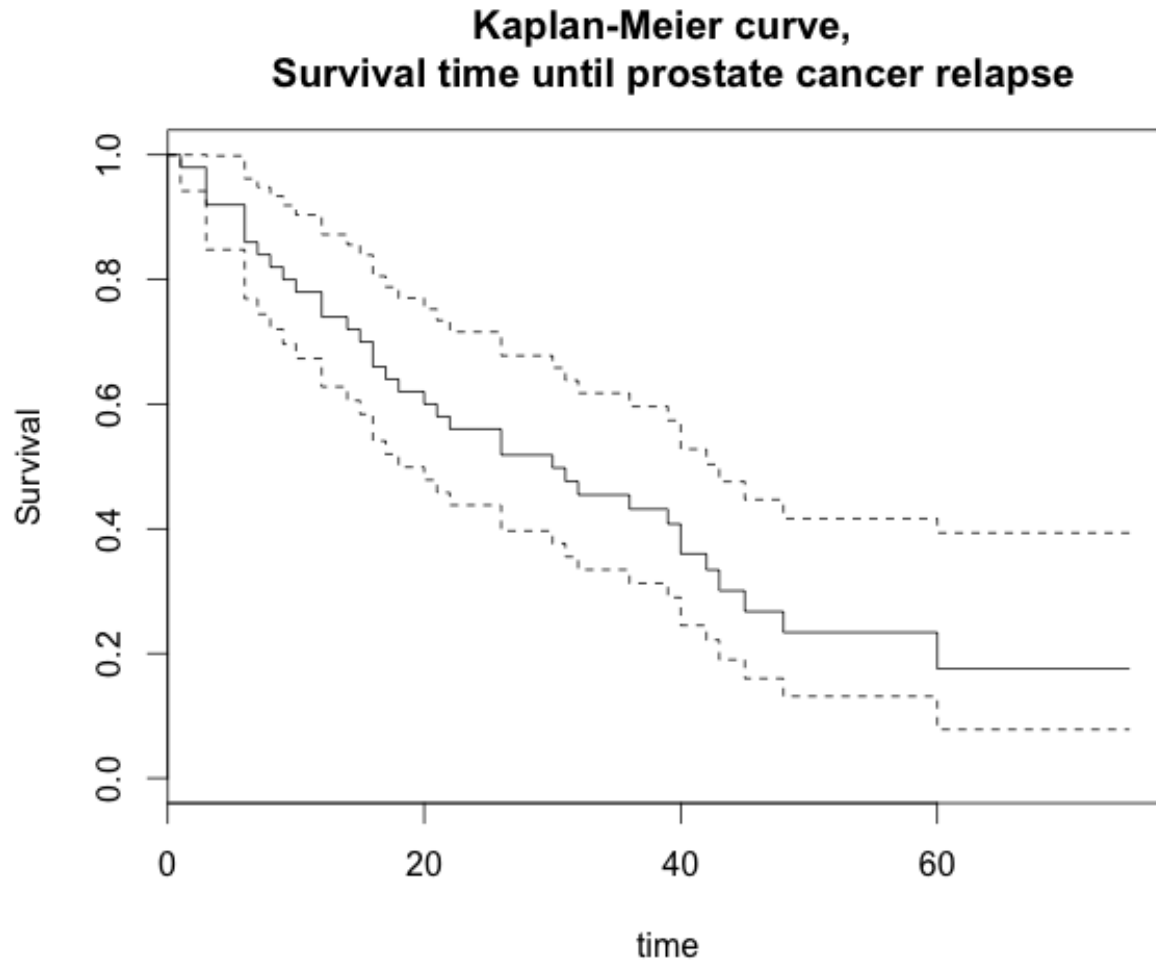
```
Call: survfit(formula = survobj ~ 1)
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	50	1	0.980	0.0198	0.9420	1.000
3	49	3	0.920	0.0384	0.8478	0.998
6	46	3	0.860	0.0491	0.7690	0.962
7	43	1	0.840	0.0518	0.7443	0.948
8	42	1	0.820	0.0543	0.7201	0.934
9	41	1	0.800	0.0566	0.6965	0.919
10	40	1	0.780	0.0586	0.6732	0.904
12	39	2	0.740	0.0620	0.6279	0.872
14	37	1	0.720	0.0635	0.6057	0.856
15	36	1	0.700	0.0648	0.5838	0.839
16	35	2	0.660	0.0670	0.5409	0.805
17	33	1	0.640	0.0679	0.5199	0.788
18	32	1	0.620	0.0686	0.4991	0.770
20	31	1	0.600	0.0693	0.4785	0.752
21	30	1	0.580	0.0698	0.4581	0.734
22	29	1	0.560	0.0702	0.4380	0.716
26	27	2	0.519	0.0709	0.3967	0.678
30	25	1	0.498	0.0710	0.3764	0.658

Estimating survival probabilities



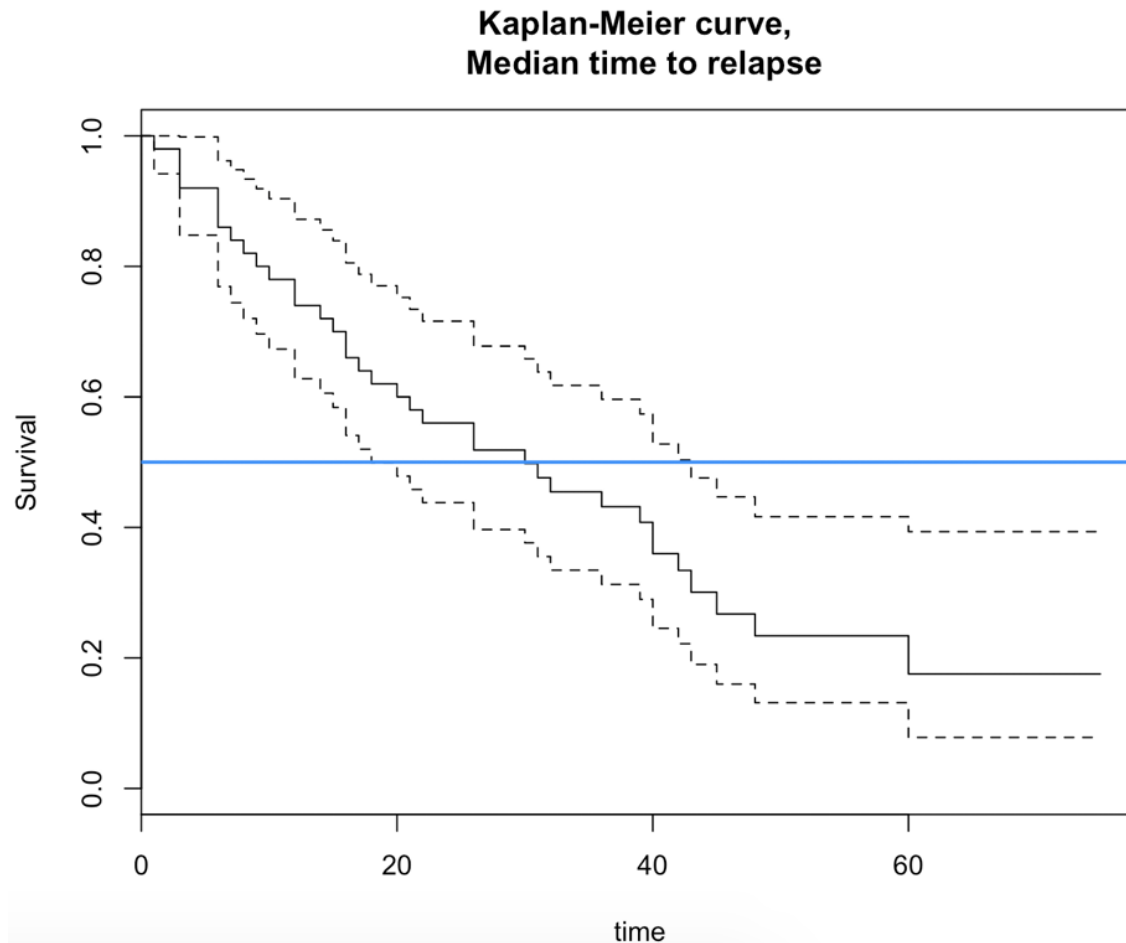
Q. What is the estimated median time until relapse?



Estimating survival probabilities



Q. What is the estimated median time until relapse?



Estimating survival probabilities



Q. What is the estimated median time until relapse?

From the output: 30 months, 95% CI: (18, 43)

```
> print(eS)
```

```
Call: survfit(formula = survobj ~ 1)
```

n	events	median	0.95LCL	0.95UCL
50	36	30	18	43

Example: PSA dataset



Suppose we are interested in determining if there is any association between time in remission and tumor grade (1-3, least to most aggressive cancer tumor).

Q. Assess this association by comparing differences in survival probabilities across tumor grade.

Stratified Kaplan-Meier



- Create a categorical tumor grade variable and obtain Kaplan-Meier survival estimates for each tumor grade group:

```
survobj<-with(psa, Surv(obstime,event))  
psa$TumorGrade <-as.factor(psa$grade)  
kms <- survfit( survobj ~ as.factor(TumorGrade), data=psa)
```

Stratified Kaplan-Meier Survival Estimates



```
> summary(kms)
```

```
Call: survfit(formula = survobj ~ as.factor(TumorGrade), data = psa)
```

9 observations deleted due to missingness

as.factor(TumorGrade)=1

time	n.risk	n.event	survival	std.err	lower	95% CI upper	95% CI
1	10	1	0.9	0.0949	0.732	1.000	
6	9	1	0.8	0.1265	0.587	1.000	
15	8	1	0.7	0.1449	0.467	1.000	
26	7	1	0.6	0.1549	0.362	0.995	
32	6	1	0.5	0.1581	0.269	0.929	
39	5	1	0.4	0.1549	0.187	0.855	
45	4	1	0.3	0.1449	0.116	0.773	
60	1	1	0.0	NaN	NA	NA	

as.factor(TumorGrade)=2

time	n.risk	n.event	survival	std.err	lower	95% CI upper	95% CI
3	15	2	0.867	0.0878	0.711	1.000	
8	13	1	0.800	0.1033	0.621	1.000	
10	12	1	0.733	0.1142	0.540	0.995	
12	11	1	0.667	0.1217	0.466	0.953	
14	10	1	0.600	0.1265	0.397	0.907	
21	9	1	0.533	0.1288	0.332	0.856	
42	6	1	0.444	0.1346	0.246	0.805	
43	4	1	0.333	0.1394	0.147	0.757	

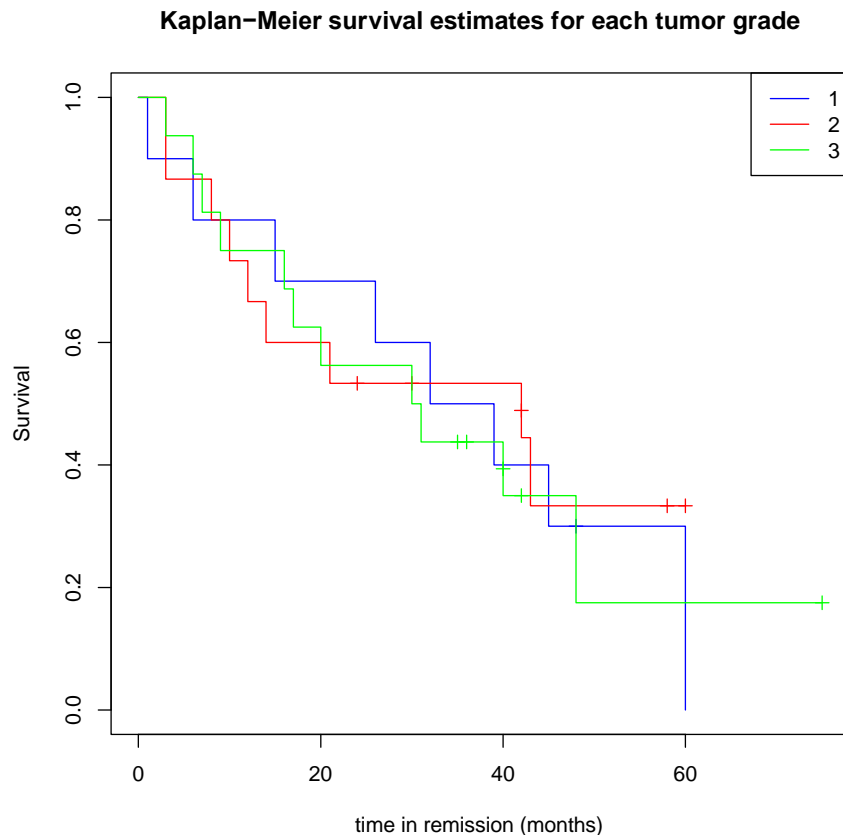
as.factor(TumorGrade)=3

time	n.risk	n.event	survival	std.err	lower	95% CI upper	95% CI
3	16	1	0.938	0.0605	0.8261	1.000	
6	15	1	0.875	0.0827	0.7271	1.000	
7	14	1	0.812	0.0976	0.6421	1.000	
9	13	1	0.750	0.1083	0.5652	0.995	
16	12	1	0.688	0.1159	0.4941	0.957	
17	11	1	0.625	0.1210	0.4276	0.914	

Stratified Kaplan-Meier Plots



```
pdf("Kaplan_Meier_Tumor_Grade_PSA.pdf")
kms <- survfit( survobj ~ TumorGrade, data=psa)
plot(kms, mark.time=TRUE,col=c("blue", "red","green"), xlab="time in remission (months)",
ylab="Survival", main="Kaplan-Meier survival estimates for each tumor grade");
legend( "topright", lty=1, col=c("blue", "red","green"), legend=levels(psa$TumorGrade))
dev.off()
```



Log rank test of differences in K-M survival estimates of tumor grades

- Perform a log rank test of differences of the Kaplan-Meier estimates of survival for two or more groups

```
> survdiff(survobj ~ TumorGrade, data=psa)
```

Call:

```
survdiff(formula = survobj ~ TumorGrade, data = psa)
```

n=41, 9 observations deleted due to missingness.

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
TumorGrade=1	10	8	7.77	0.00681	0.00963
TumorGrade=2	15	9	10.13	0.12576	0.20014
TumorGrade=3	16	11	10.10	0.07992	0.12841

Chisq= 0.2 on 2 degrees of freedom, p= 0.9