

Companion Code and Output for Lecture Slides on
Preliminaries/Kaplan-Meier
B. Jones, Dept. of Political Science
UC-Davis

Kaplan-Meier Estimate of Survivor Function

Using the U.N. Peacekeeping Mission data (see my website for access to this data).
Stata code and output (data was already `stset`):

```
. sts list

      failure _d: failed
analysis time _t: duration
```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf. Int.]
2	54	1	0	0.9815	0.0183	0.8757 0.9974
4	53	1	0	0.9630	0.0257	0.8599 0.9906
5	52	2	0	0.9259	0.0356	0.8146 0.9715
7	50	3	0	0.8704	0.0457	0.7472 0.9360
10	47	1	0	0.8519	0.0483	0.7255 0.9230
11	46	1	0	0.8333	0.0507	0.7042 0.9096
12	45	2	0	0.7963	0.0548	0.6624 0.8816
13	43	1	1	0.7778	0.0566	0.6420 0.8672
15	41	2	0	0.7398	0.0598	0.6005 0.8369
16	39	1	1	0.7209	0.0612	0.5802 0.8214
18	37	1	0	0.7014	0.0626	0.5594 0.8053
19	36	1	0	0.6819	0.0638	0.5389 0.7889
21	35	1	0	0.6624	0.0649	0.5186 0.7723
23	34	2	1	0.6235	0.0667	0.4789 0.7385
25	31	3	1	0.5631	0.0687	0.4185 0.6848
26	27	1	0	0.5423	0.0693	0.3980 0.6660
28	26	1	0	0.5214	0.0697	0.3777 0.6469
29	25	1	1	0.5005	0.0699	0.3577 0.6276
30	23	2	0	0.4570	0.0703	0.3164 0.5870<---Taken to be Median S(t)
31	21	1	0	0.4353	0.0702	0.2962 0.5663
34	20	1	0	0.4135	0.0700	0.2764 0.5453
46	19	3	0	0.3482	0.0684	0.2189 0.4807
48	16	2	0	0.3047	0.0664	0.1823 0.4362
49	14	1	0	0.2829	0.0651	0.1645 0.4134
59	13	0	1	0.2829	0.0651	0.1645 0.4134
66	12	1	0	0.2593	0.0638	0.1453 0.3890
70	11	1	0	0.2358	0.0622	0.1266 0.3642
71	10	0	1	0.2358	0.0622	0.1266 0.3642
99	9	0	1	0.2358	0.0622	0.1266 0.3642
127	8	0	2	0.2358	0.0622	0.1266 0.3642
128	6	1	0	0.1965	0.0630	0.0912 0.3310
284	5	0	1	0.1965	0.0630	0.0912 0.3310
319	4	0	1	0.1965	0.0630	0.0912 0.3310
452	3	0	1	0.1965	0.0630	0.0912 0.3310
634	2	0	1	0.1965	0.0630	0.0912 0.3310
641	1	0	1	0.1965	0.0630	0.0912 0.3310

R code and output (use survival library; foreign is used to import Stata dataset:

```
> KM<-survfit(Surv(duration, failed), data = undata)
```

```
>
```

```
> summary(KM)
```

```
Call: survfit(formula = Surv(duration, failed), data = undata)
```

4 observations deleted due to missingness

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
2	54	1	0.981	0.0183	0.946	1.000
4	53	1	0.963	0.0257	0.914	1.000
5	52	2	0.926	0.0356	0.859	0.998
7	50	3	0.870	0.0457	0.785	0.965
10	47	1	0.852	0.0483	0.762	0.952
11	46	1	0.833	0.0507	0.740	0.939
12	45	2	0.796	0.0548	0.696	0.911
13	43	1	0.778	0.0566	0.674	0.897
15	41	2	0.740	0.0598	0.631	0.867
16	39	1	0.721	0.0612	0.610	0.851
18	37	1	0.701	0.0626	0.589	0.835
19	36	1	0.682	0.0638	0.568	0.819
21	35	1	0.662	0.0649	0.547	0.803
23	34	2	0.623	0.0667	0.506	0.769
25	31	3	0.563	0.0687	0.443	0.715
26	27	1	0.542	0.0693	0.422	0.697
28	26	1	0.521	0.0697	0.401	0.678
29	25	1	0.501	0.0699	0.381	0.658
30	23	2	0.457	0.0703	0.338	0.618
31	21	1	0.435	0.0702	0.317	0.597
34	20	1	0.413	0.0700	0.297	0.576
46	19	3	0.348	0.0684	0.237	0.512
48	16	2	0.305	0.0664	0.199	0.467
49	14	1	0.283	0.0651	0.180	0.444
66	12	1	0.259	0.0638	0.160	0.420
70	11	1	0.236	0.0622	0.141	0.395
128	6	1	0.196	0.0630	0.105	0.369

-----Taken to
be Median S(t)

It is often useful to plot Kaplan-Meier estimates:

Stata

The command `sts graph` returns (I added labels, etc. on my own, code not shown):

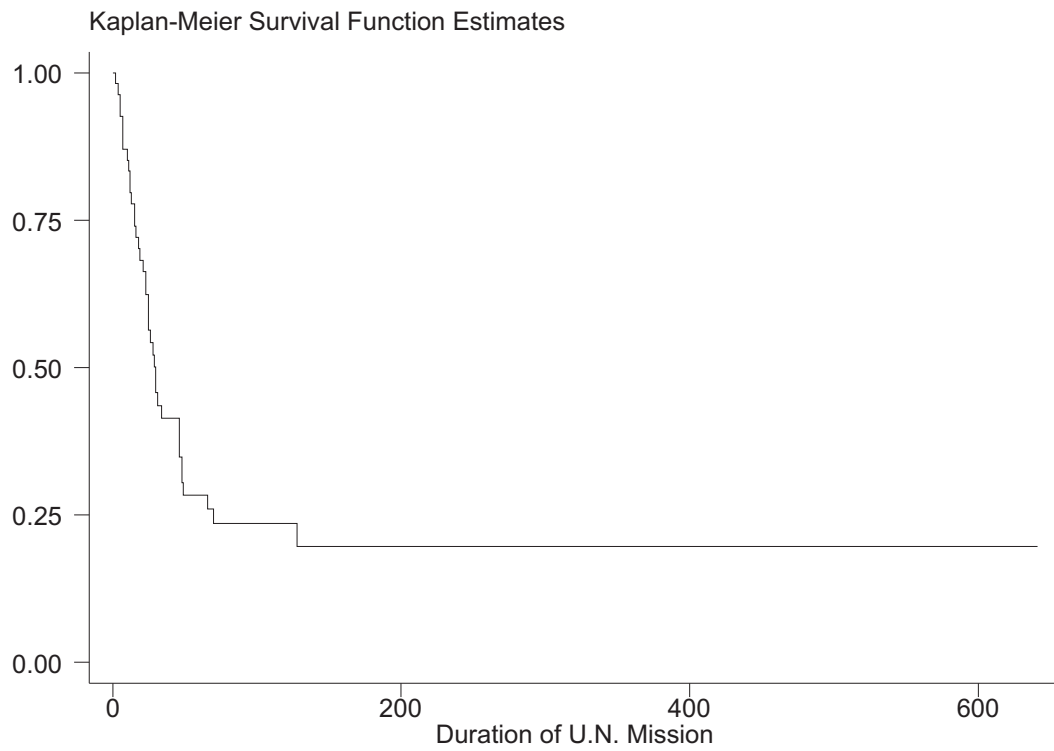


Figure 1: *This graph plots the nonparametric Kaplan-Meier survival function for the U.N. Peacekeeping Mission data. Note the stair-step nature of the figure.*

R

```
plotUN<-plot(survfit(Surv(duration, failed), data = undata),  
  main = "Kaplan-Meier Estimates",  
  ylab= "Probability" ,  
  xlab= "Survival Time in Months")
```

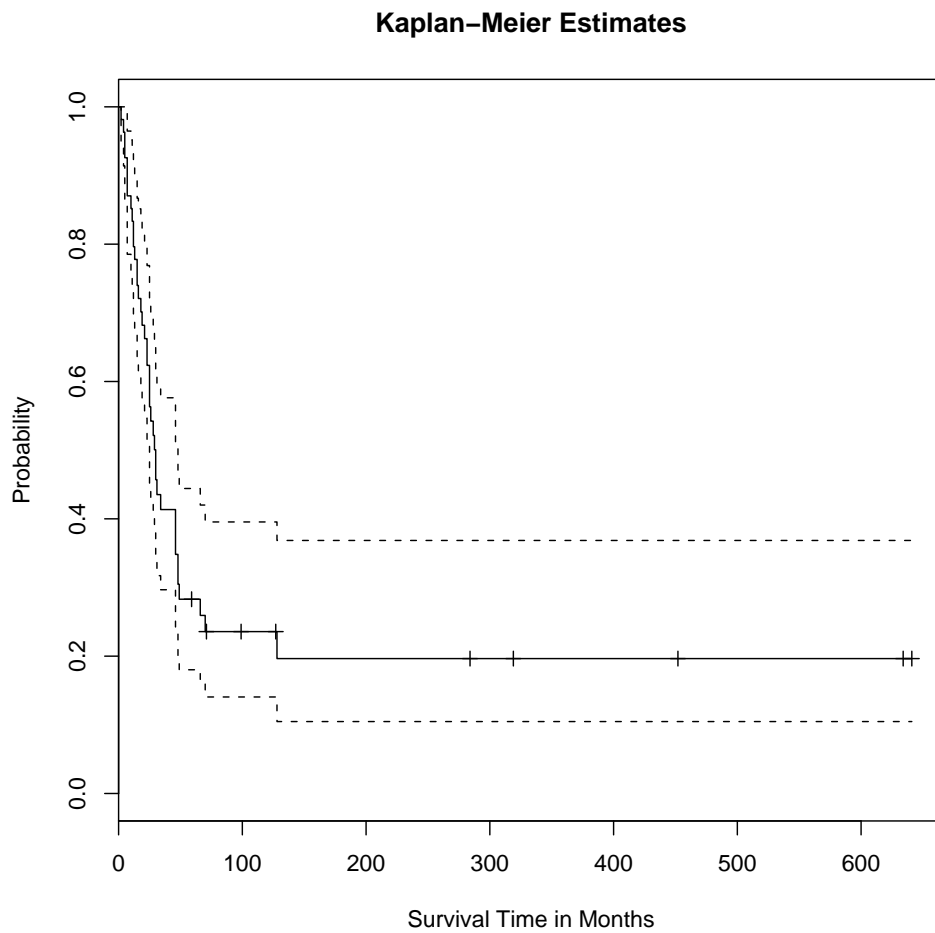


Figure 2: *This graph plots the nonparametric Kaplan-Meier survival function for the U.N. Peacekeeping Mission data. Note the stair-step nature of the figure.*

Kaplan-Meier estimates may gotten across subgroups:

```
> KMtype<-survfit(Surv(duration, failed)~mistype, data = undata)
>
> summary(KMtype)
Call: survfit(formula = Surv(duration, failed) ~ mistype, data = undata)
```

```
12 observations deleted due to missingness
mistype=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 2   24      1   0.958  0.0408  0.8816  1.000
 5   23      2   0.875  0.0675  0.7522  1.000
 7   21      2   0.792  0.0829  0.6448  0.972
15   19      1   0.750  0.0884  0.5953  0.945
16   18      1   0.708  0.0928  0.5480  0.916
18   17      1   0.667  0.0962  0.5024  0.885
21   16      1   0.625  0.0988  0.4585  0.852
23   15      1   0.583  0.1006  0.4160  0.818
26   14      1   0.542  0.1017  0.3749  0.783
29   13      1   0.500  0.1021  0.3351  0.746
30   12      1   0.458  0.1017  0.2967  0.708
31   11      1   0.417  0.1006  0.2595  0.669
46   10      2   0.333  0.0962  0.1893  0.587
48    8      1   0.292  0.0928  0.1564  0.544
49    7      1   0.250  0.0884  0.1250  0.500
66    5      1   0.200  0.0837  0.0881  0.454

mistype=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 7    7      1   0.857  0.132  0.633  1
48    6      1   0.714  0.171  0.447  1
70    5      1   0.571  0.187  0.301  1
128   4      1   0.429  0.187  0.182  1

mistype=3
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 4   15      1   0.933  0.0644  0.8153  1.000
11   14      1   0.867  0.0878  0.7106  1.000
12   13      2   0.733  0.1142  0.5405  0.995
13   11      1   0.667  0.1217  0.4661  0.953
19   10      1   0.600  0.1265  0.3969  0.907
25    9      3   0.400  0.1265  0.2152  0.743
28    6      1   0.333  0.1217  0.1630  0.682
30    5      1   0.267  0.1142  0.1152  0.617
34    4      1   0.200  0.1033  0.0727  0.550
46    3      1   0.133  0.0878  0.0367  0.484
```

I could plot these via (output not given here):

```
plotUN<-plot(survfit(Surv(duration, failed)~mistype, data = undata),
  main = "Kaplan-Meier Estimates",
  ylab= "Probability" ,
  xlab= "Survival Time in Months")
```