

Dataset

veteran: dataset of 137 observations x 8 variables from a two-treatment randomized trial for lung cancer.

trt: 1=standard 2=test

celltype: 1=squamous, 2=smallcell, 3=adeno, 4=large

time: survival time

status: censoring status

karno: Karnofsky performance score (100=good)

diagtime: months from diagnosis to randomisation

ageA: in years

prior: prior therapy 0=no, 1=yes

```
1 > head(veteran)
2   trt celltype time status karno diagtime age prior
3 1    1 squamous  72      1    60        7  69     0
4 2    1 squamous 411      1    70        5  64    10
5 3    1 squamous 228      1    60        3  38     0
6 4    1 squamous 126      1    60        9  63    10
7 5    1 squamous 118      1    70       11  65    10
8 6    1 squamous  10      1    20        5  49     0
```

Summary

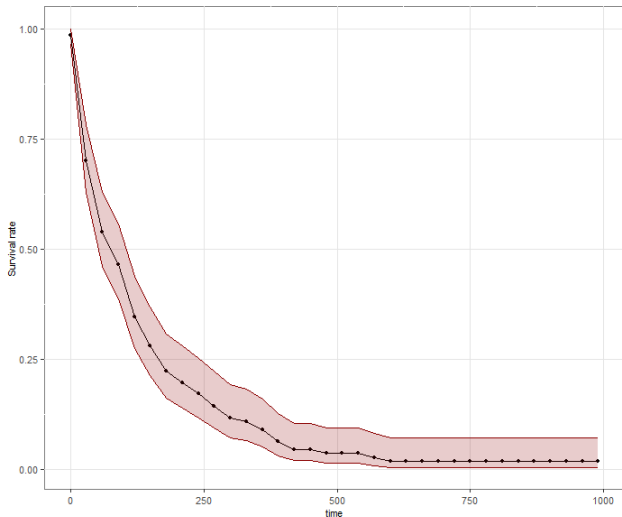
From this initial summary which has class "summary survfit", we will make a dataframe ready for plotting withing the ggplot2 environment.

```
1 > survival.30.dataset = summary(kma_1, times = c(1, (1:33)*30))
2 > survival.30.dataset
3 Call: survfit(formula = Surv(time, status) ~ 1, data = veteran)
4
5   time  n.risk  n.event  survival  std.err  lower 95% CI  upper 95% CI
6     1      137        2    0.985  0.0102    0.96552  1.0000
7     30       97       39    0.700  0.0392    0.62774  0.7816
8     60       73       22    0.538  0.0427    0.46070  0.6288
9     90       62       10    0.464  0.0428    0.38731  0.5560
10    120       43       15    0.346  0.0414    0.27345  0.4372
11    150       34        8    0.280  0.0395    0.21240  0.3693
12    180       27        7    0.222  0.0369    0.16066  0.3079
13    210       23        3    0.197  0.0355    0.13814  0.2802
14    240       19        3    0.171  0.0338    0.11613  0.2520
15    270       16        3    0.144  0.0319    0.09338  0.2223
16    300       13        3    0.117  0.0295    0.07147  0.1917
17    330       12        1    0.108  0.0285    0.06439  0.1813
18    360       10        2    0.090  0.0265    0.05061  0.1602
19 ...
```

Kaplan-Meier analysis

Kaplan-Meier analysis

Veteran data



	time	surv
1	1	0.9854
2	30	0.7004
3	60	0.5382
4	90	0.464
5	120	0.3458
6	150	0.2801
7	180	0.2224
8	210	0.1967
9	240	0.1711
10	270	0.1441
11	300	0.1171
12	330	0.1081
13	360	0.09
14	390	0.063
15	420	0.045
16	450	0.045
17	480	0.036
18	510	0.036
19	540	0.036
20	570	0.027

Main observations

- At the first month (after 30 days), the survival rate or probability of survival is about 70%.
- There seems to be some kind of breakup point at 6 months (after 180 days) as the slope gets less steep.
- After one year, the survival rate is lower than 10%. A patient has a 10% or less probability of surviving one year.

R code (1/4)

```
1 > #load libraries
2 > library(survival)
3 > library(ggplot2)
4 > library(gridExtra)
5
6 > # load data and head of the dataset
7 > data(veteran)
8 > head(veteran)
```

	trt	celltype	time	status	karno	diagtime	age	prior
10 1	1	squamous	72	1	60	7	69	0
11 2	1	squamous	411	1	70	5	64	10
12 3	1	squamous	228	1	60	3	38	0
13 4	1	squamous	126	1	60	9	63	10
14 5	1	squamous	118	1	70	11	65	10
15 6	1	squamous	10	1	20	5	49	0

R code (2/4)

```
1 > km = with(veteran, Surv(time, status))
2 > head(km, 100)
3   [1] 72 411 228 126 118 10 82 110 314 100+ 42 8 144 25+ 11
      30 384 4 54 13 123+
4  [22] 97+ 153 59 117 16 151 22 56 21 18 139 20 31 52 287
      18 51 122 27 54 7
5  [43] 63 392 10 8 92 35 117 132 12 162 3 95 177 162 216
      553 278 12 260 200 156
6  [64] 182+ 143 105 103 250 100 999 112 87+ 231+ 242 991 111 1 587
      389 33 25 357 467 201
7  [85] 1 30 44 283 15 25 103+ 21 13 87 2 20 7 24 99
      8
```

R code 3/4)

```
1 > # Kaplan-Meier estimates of the probability of survival over time
2 > kma_1 = survfit(Surv(time, status) ~ 1, data=veteran)
3 > # max time: 999 days (about 33 months (30 days))
4 > survival.30.dataset = summary(kma_1, times = c(1, (1:33)*30))
5 > # convert summary to data.frame for plotting
6 > cols = lapply(1:15 , function(x) survival.30.dataset[x])
7 > df = do.call(data.frame, cols)
8 >
9 > # table to be displayed next to the graph as a second graph
10 > df2 = df[1:20, c(2,6)]
11 > df2$surv = round(df2$surv, 4)
```

R code (4/4)

```
1 # KM plot (ggplot2)
2 p1 = ggplot(df, aes(x = time, y = surv)) +
3   geom_line(color = 'black') +
4   geom_point(size = 1.2) +
5   geom_ribbon(aes(ymin = lower, ymax = upper), alpha=0.2, fill= 'darkred', col =
      'darkred') +
6
7   labs(title = 'Kaplan-Meier analysis',
8         subtitle = 'Veteran data',
9         y="Survival rate", x="time") +
10  theme(axis.text=element_text(size=8),
11        axis.title=element_text(size=8),
12        plot.subtitle=element_text(size=9, face="italic", color="darkred"),
13        panel.background = element_rect(fill = "white", colour = "grey50"),
14        panel.grid.major = element_line(colour = "grey90"))
15
16 p2 = tableGrob(df2)
17
18 grid.arrange(p1, p2, ncol = 2, nrow = 1, widths = c(6, 2))
```