

# 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

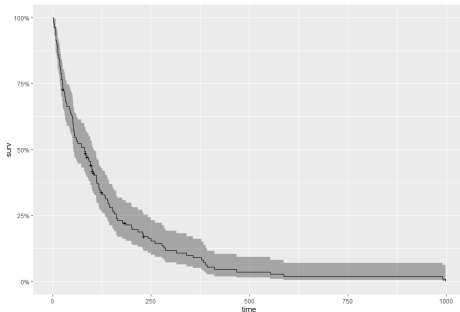
**prior:** prior therapy 0=no, 1=yes

```
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
```

# Kaplan-Meier estimator

Kaplan-Meier estimate of the probability of survival over time and plotting.

```
1 km_fit = survfit(Surv(time, status) ~ 1, data=veteran)
2 summary(km_fit, times = c(1,30,60,90*(1:10)))
3 time n.risk n.event survival std.err lower 95% CI upper 95% CI
4 1 137 2 0.985 0.0102 0.96552 1.0000
5 30 97 39 0.700 0.0392 0.62774 0.7816
6 60 73 22 0.538 0.0427 0.46070 0.6288
7 90 62 10 0.464 0.0428 0.38731 0.5560
8 180 27 30 0.222 0.0369 0.16066 0.3079
9 ...
10 autoplot(km_fit)
```



# Kaplan-Meier estimator by treatment

Kaplan-Meier estimate of the probability of survival over time by treatment and for patients over 60.

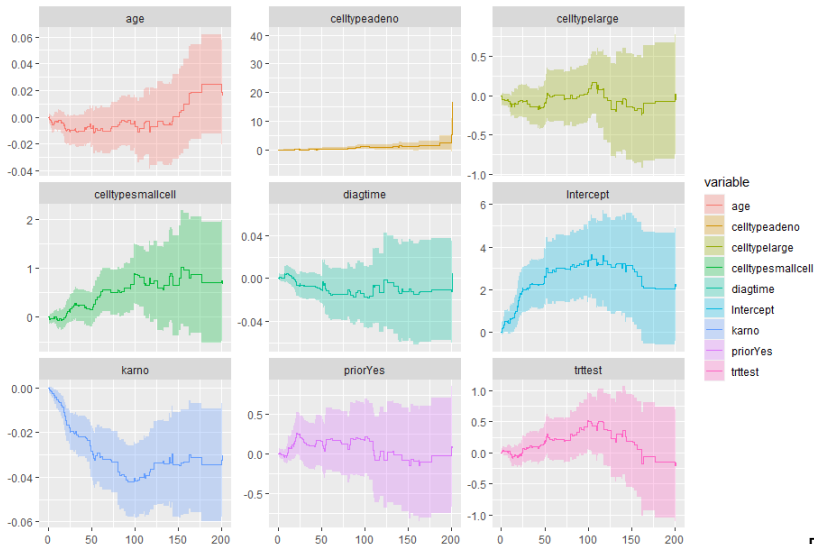
```
1 vet = mutate(veteran, AG = ifelse((age < 60), "LT60", "OV60"),  
2             AG = factor(AG),  
3             trt = factor(trt, labels=c("standard", "test")),  
4             prior = factor(prior, labels=c("NO", "Yes")))  
5  
6 km_AG_fit = survfit(Surv(time, status) ~ AG, data=vet)  
7 autoplot(km_AG_fit)
```



# Cox proportional hazard model

```
1 > cox <- coxph(Surv(time, status) ~ trt + celltype + karno + diagtime + age +
  prior, data = vet)
2 > summary(cox)
3
4 coxph(formula = Surv(time, status) ~ trt + celltype + karno +
  diagtime + age + prior, data = vet)
5
6
7              coef exp(coef)    se(coef)      z Pr(>|z|)
8 trttest          2.946e-01  1.343e+00  2.075e-01  1.419  0.15577
9 celltypesmallcell 8.616e-01  2.367e+00  2.753e-01  3.130  0.00175 **
10 celltypeadeno    1.196e+00  3.307e+00  3.009e-01  3.975  7.05e-05 ***
11 celltypelarge    4.013e-01  1.494e+00  2.827e-01  1.420  0.15574
12 karno           -3.282e-02  9.677e-01  5.508e-03 -5.958  2.55e-09 ***
13 diagtime         8.132e-05  1.000e+00  9.136e-03  0.009  0.99290
14 age             -8.706e-03  9.913e-01  9.300e-03 -0.936  0.34920
15 priorYes         7.159e-02  1.074e+00  2.323e-01  0.308  0.75794
16 ---
17
18              exp(coef) exp(-coef) lower .95 upper .95
19 trttest          1.3426      0.7448    0.8939    2.0166
20 celltypesmallcell 2.3669      0.4225    1.3799    4.0597
21 celltypeadeno    3.3071      0.3024    1.8336    5.9647
22 celltypelarge    1.4938      0.6695    0.8583    2.5996
23 karno            0.9677      1.0334    0.9573    0.9782
24 diagtime         1.0001      0.9999    0.9823    1.0182
25 age              0.9913      1.0087    0.9734    1.0096
26 priorYes         1.0742      0.9309    0.6813    1.6937
27
28 Concordance= 0.736 (se = 0.021 )
29 Likelihood ratio test= 62.1 on 8 df, p=2e-10
30 Wald test            = 62.37 on 8 df, p=2e-10
31 Score (logrank) test = 66.74 on 8 df, p=2e-11
```

# Aalen's additive regression model for censored data



# Random forest model (1/2)

```
1 # ranger model
2 r_fit <- ranger(Surv(time, status) ~ trt + celltype +
3               karno + diagtime + age + prior,
4               data = vet,
5               mtry = 4,
6               importance = "permutation",
7               splitrule = "extratrees",
8               verbose = TRUE)
9
10 # Average the survival models
11 death_times <- r_fit$unique.death.times
12 surv_prob <- data.frame(r_fit$survival)
13 avg_prob <- sapply(surv_prob, mean)
14
15 # Plot the survival models for each patient
16 plot(r_fit$unique.death.times, r_fit$survival[1,], type = "l", ylim = c(0,1),
17      col = "red", xlab = "Days", ylab = "survival", main = "Patient Survival
18      Curves")
19 cols <- colors()
20 for (n in sample(c(2:dim(vet)[1]), 20)){
21   lines(r_fit$unique.death.times, r_fit$survival[n,], type = "l", col = cols[n])
22 }
23 lines(death_times, avg_prob, lwd = 2)
24 legend(500, 0.7, legend = c('Average = black'))
```

# Random forest model (2/2)

