Nicotine Addiction

Stratified logrank test: pharmacoSmoking dataset

1. Show the KM
2. Compute p-value

```
library(asaur)
library(survival)
```

```
##
## Attaching package: 'survival'

## The following object is masked from 'package:rpart':
##
##     solder
```

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------- tidy

## v ggplot2 3.1.0     v purrr   0.3.0
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0

## -- Conflicts ------------------------------------------------------------------------- tidyverse_
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

The data

```
dat <- pharmacoSmoking
head(dat)
```

```
##    id ttr relapse         grp age gender     race employment yearsSmoking
## 1  21 182       0   patchOnly  36   Male    white         ft           26
## 2 113  14       1   patchOnly  41   Male    white      other           27
## 3  39   5       1 combination  25 Female    white      other           12
## 4  80  16       1 combination  54   Male    white         ft           39
## 5  87   0       1 combination  45   Male    white      other           30
## 6  29 182       0 combination  43   Male hispanic         ft           30
##   levelSmoking ageGroup2 ageGroup4 priorAttempts longestNoSmoke
## 1        heavy     21-49     35-49             0              0
## 2        heavy     21-49     35-49             3             90
## 3        heavy     21-49     21-34             3             21
## 4        heavy       50+     50-64             0              0
## 5        heavy     21-49     35-49             0              0
## 6        heavy     21-49     35-49             2           1825
```

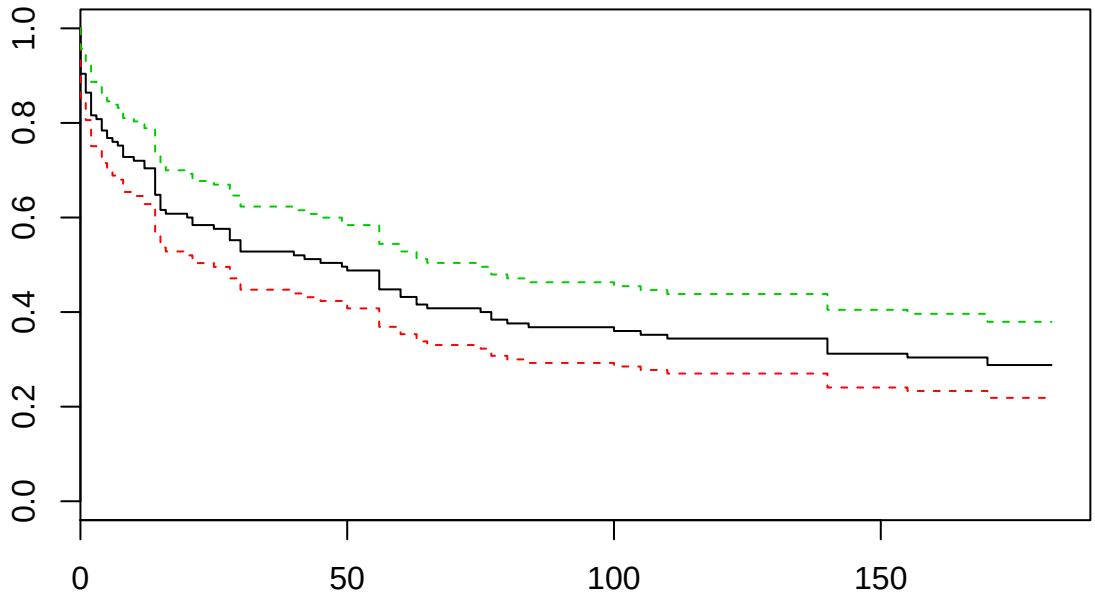
```
summary(dat)
```

```
##        id              ttr            relapse              grp
##  Min.   :  1.00   Min.   :  0.00   Min.   :0.000   combination:61
##  1st Qu.: 33.00   1st Qu.:  8.00   1st Qu.:0.000   patchOnly  :64
##  Median : 67.00   Median : 49.00   Median :1.000
##  Mean   : 66.15   Mean   : 77.44   Mean   :0.712
##  3rd Qu.: 99.00   3rd Qu.:182.00   3rd Qu.:1.000
##  Max.   :130.00   Max.   :182.00   Max.   :1.000
##       age           gender         race    employment  yearsSmoking
##  Min.   :22.00   Female:81   black   :38   ft   :72   Min.   : 9.00
##  1st Qu.:41.00   Male  :44   hispanic: 8   other:39   1st Qu.:22.00
##  Median :49.00               other   : 2   pt   :14   Median :30.00
##  Mean   :48.84               white   :77              Mean   :30.88
##  3rd Qu.:56.00                                       3rd Qu.:39.00
##  Max.   :86.00                                       Max.   :56.00
##  levelSmoking ageGroup2  ageGroup4  priorAttempts     longestNoSmoke
##  heavy:89     21-49:66   21-34:16   Min.   :   0.00   Min.   :   0.0
##  light:36     50+  :59   35-49:50   1st Qu.:   1.00   1st Qu.:   7.0
##                          50-64:48   Median :   2.00   Median :  90.0
##                          65+  :11   Mean   :  12.68   Mean   : 539.7
##                                     3rd Qu.:   5.00   3rd Qu.: 365.0
##                                     Max.   :1000.00   Max.   :6205.0
```

```
fit.KM <- survfit(Surv(ttr, relapse) ~ 1, data = dat)
fit.KM
```

```
## Call: survfit(formula = Surv(ttr, relapse) ~ 1, data = dat)
##
##       n  events  median 0.95LCL 0.95UCL
##     125      89      49      25      75
```

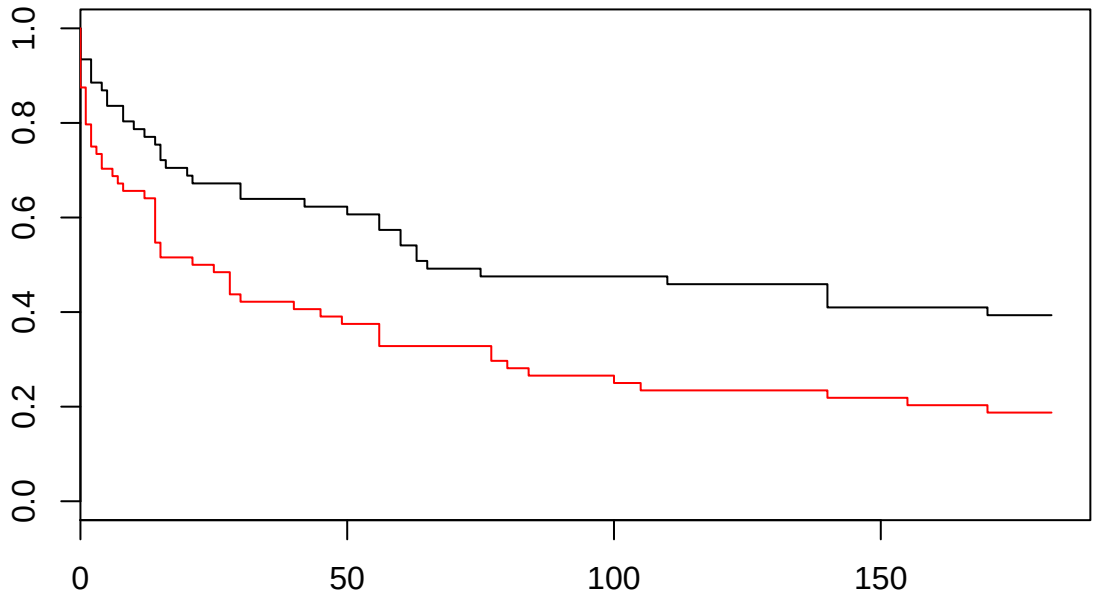
```
plot(fit.KM, col = 1:3) #col1:black col2:red col3:green
```




```
fit.KM <- survfit(Surv(ttr, relapse) ~ grp, data = dat)
fit.KM #col1:combination col2:patchOnly
```

```
## Call: survfit(formula = Surv(ttr, relapse) ~ grp, data = dat)
##
##                   n events median 0.95LCL 0.95UCL
## grp=combination 61     37     65      50      NA
## grp=patchOnly   64     52     23      14      56
```

```
plot(fit.KM, col = 1:2) #col1:black col2:red col3:green
```



Question: do the 2 treatment group differ significantly in terms of survival to relapse?

```
survdiff(Surv(ttr, relapse) ~ grp, data = dat)
```

```
## Call:
## survdiff(formula = Surv(ttr, relapse) ~ grp, data = dat)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## grp=combination 61       37     49.9      3.36      8.03
## grp=patchOnly   64       52     39.1      4.29      8.03
##
##  Chisq= 8  on 1 degrees of freedom, p= 0.005
```

Combination therapy is significantly better!

Critique: the 2 groups have different age distribution, which might confound our results. Lets investigate:

```
with(dat, prop.table(table(grp, ageGroup2), 1))
```

```
##              ageGroup2
## grp               21-49       50+
##   combination 0.5081967 0.4918033
##   patchOnly   0.5468750 0.4531250
```

stratified logrank test

```
survdiff(Surv(ttr, relapse) ~ grp + strata(ageGroup2), data = dat)
```

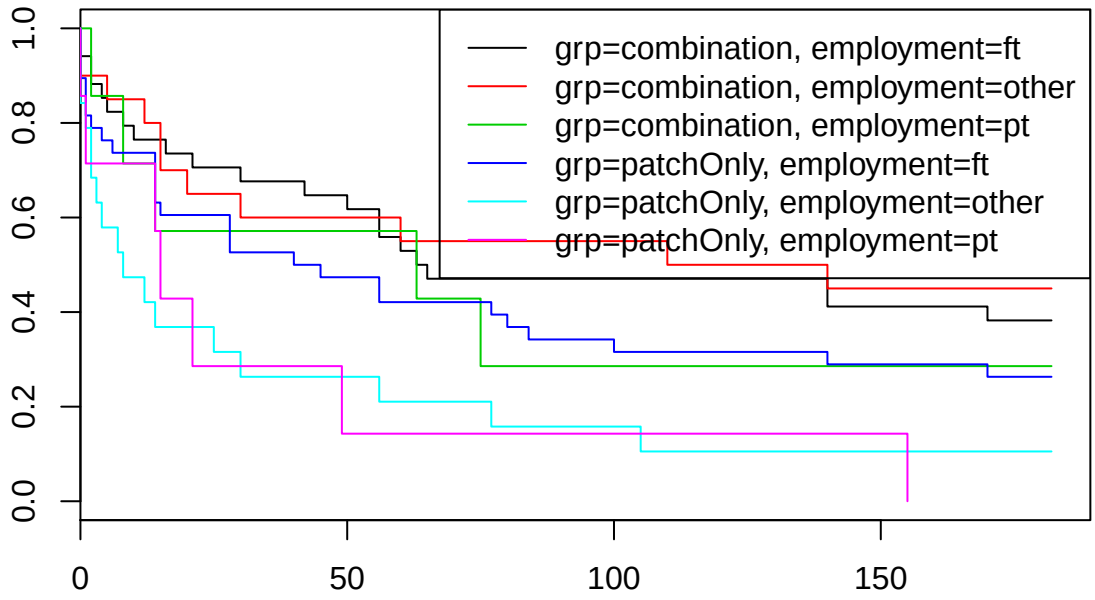
```
## Call:
## survdiff(formula = Surv(ttr, relapse) ~ grp + strata(ageGroup2),
##     data = dat)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## grp=combination 61       37     49.1      2.99      7.03
## grp=patchOnly   64       52     39.9      3.68      7.03
##
##  Chisq= 7  on 1 degrees of freedom, p= 0.008
```

extra

```
fit.4 <- survfit(Surv(ttr, relapse) ~ grp + employment, data = dat)
fit.4
```

```
## Call: survfit(formula = Surv(ttr, relapse) ~ grp + employment, data = dat)
##
##                                    n events median 0.95LCL 0.95UCL
## grp=combination, employment=ft    34     21   64.0      50      NA
## grp=combination, employment=other 20     11  125.0      20      NA
## grp=combination, employment=pt     7      5   63.0       8      NA
## grp=patchOnly, employment=ft      38     28   42.5      14     140
## grp=patchOnly, employment=other   19     17    8.0       3      77
## grp=patchOnly, employment=pt       7      7   15.0       1      NA
```

```
plot(fit.4, col = 1:6)
legend("topright", lty = 1, col = 1:6, legend = names(fit.4$strata))
```



The 3 'combination' curves seem all higher than the 3 'patchOnly' curves. Lets make a stratified test:

```
survdiff(Surv(ttr, relapse) ~ grp + strata(employment), data = dat)
```

```
## Call:
## survdiff(formula = Surv(ttr, relapse) ~ grp + strata(employment),
##     data = dat)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## grp=combination 61       37     50.3      3.50      8.58
## grp=patchOnly   64       52     38.7      4.54      8.58
##
##  Chisq= 8.6  on 1 degrees of freedom, p= 0.003
```