Practical_survival_1_R.R

classale

Tue May 10 15:14:55 2016

```
#################################################################
## load packages                                               ##
#################################################################
#install.packages("survival")
#install.packages("ggplot2")
library(survival)
```

```
## Loading required package: splines
```

```
library(ggplot2)
#################################################################
## 1. Set working directory and get data                       ##
#################################################################
setwd("C:/Users/classale/STATS/")
load("elsa_cf.rdata")
attach(elsa_cf)

#################################################################
## 2. Getting a sense of the data                              ##
#################################################################
# Get list of variables in the dataset 'elsa_cf'
names(elsa_cf)
```

```
##  [1] "cf1"        "cancer1"    "cigst1"     "sex"        "age1"
##  [6] "dodmnth"    "dodyr"      "educ1"      "chd1"       "totwq5_bu1"
## [11] "iintdtm1"   "iintdty1"   "death"      "alcohol1"   "physinact1"
## [16] "time"       "id"
```

```
# Get summary statistic for 'age1' and 'time'
summary(elsa_cf$age1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   50.00   57.00   63.00   64.31   71.00   90.00
```

```
summary(elsa_cf$time)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  0.04107 10.20000 10.62000  9.60500 10.79000 11.04000
```

```
# Histogram separately for people with an event and survivors #

#Create a subset#
elsadead <-subset(elsa_cf, elsa_cf$death==1)
elsalive <-subset(elsa_cf, elsa_cf$death==0)

summary(elsadead$time)
```

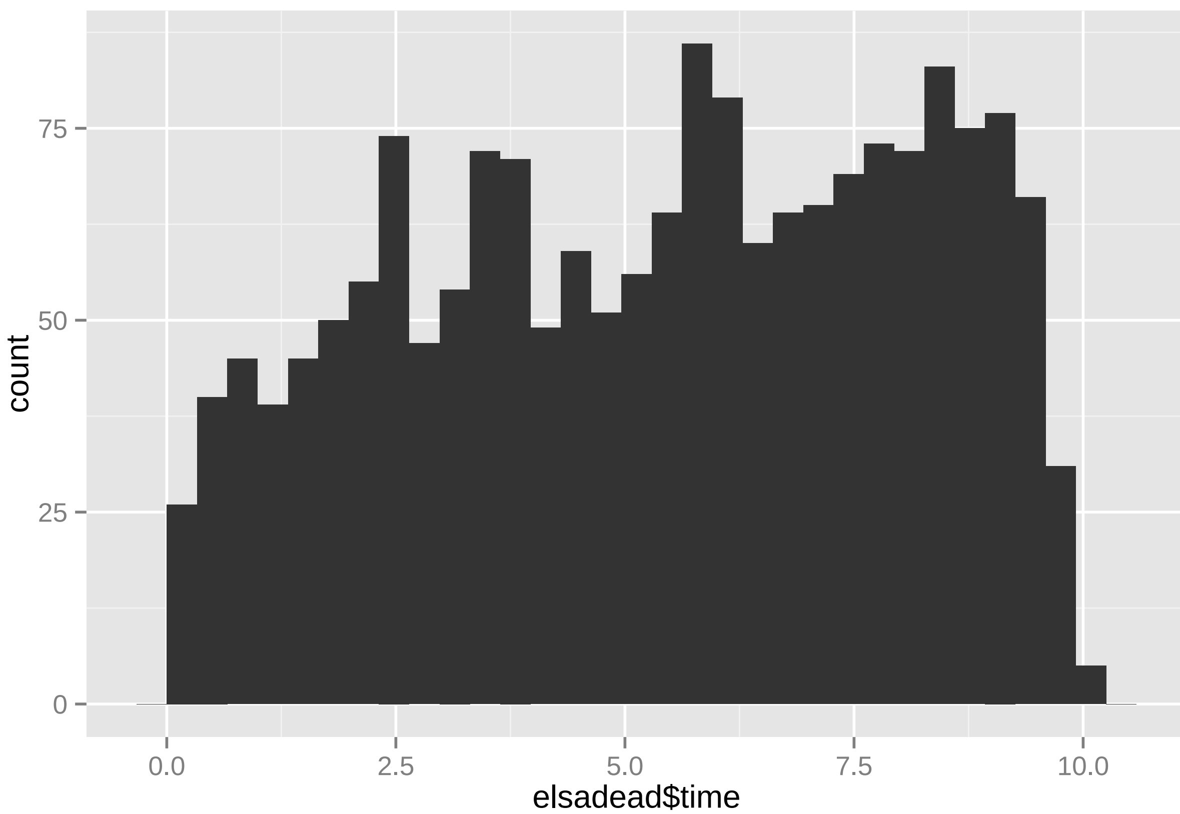
```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.04107 3.20500 5.62400 5.38000 7.70700 9.96300
```

```r
summary(elsalive$time)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.04   10.46   10.71   10.62   10.79   11.04
```

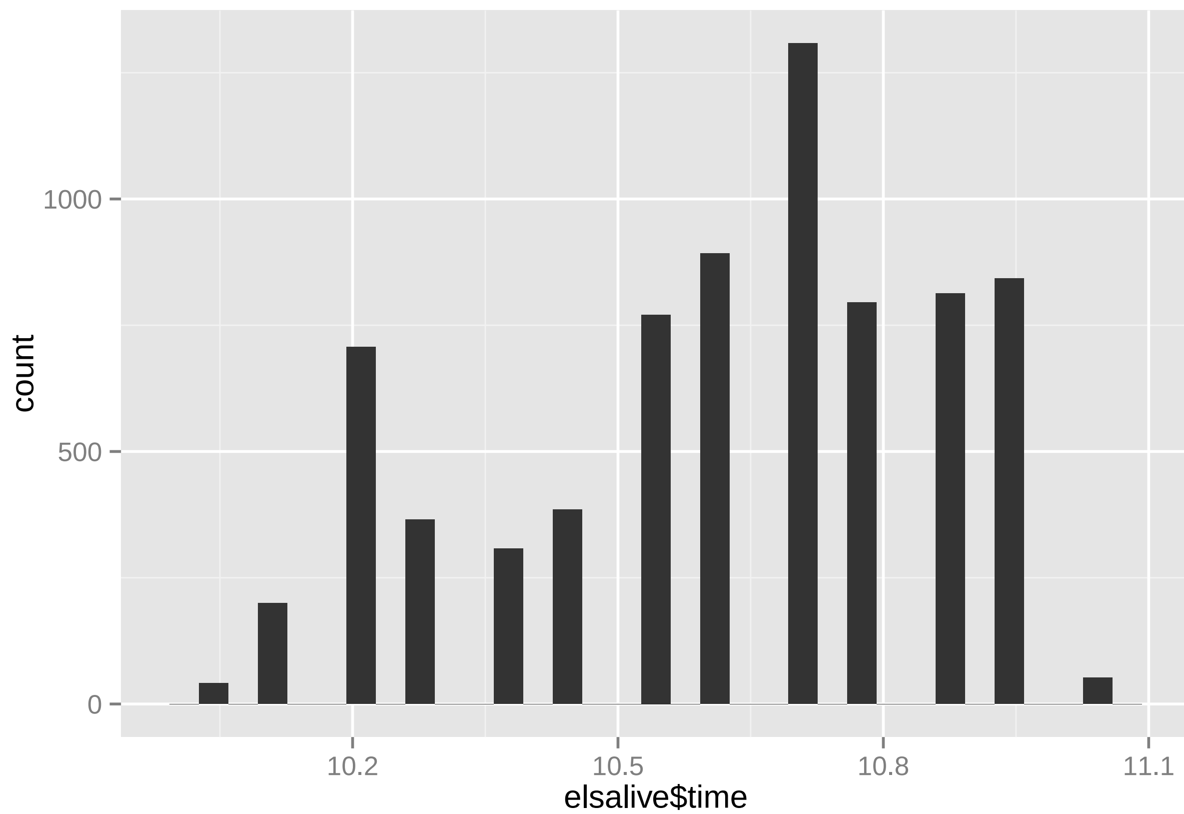
```r
# Histogram in each subset#
qplot(elsadead$time, geom = 'histogram', bins = 40)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```




```r
qplot(elsalive$time, geom = 'histogram', bins = 40)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```




```
# Tabulate the outcome 'death'
table(elsa_cf$death)
```

```
##
##    0    1
## 7480 1802
```

```
prop.table(table(elsa_cf$death))
```

```
##
##         0         1
## 0.8058608 0.1941392
```

```
##############################################################
## 3. Survival and KM curve                                 ##
##############################################################
# Calculate person-years and incidence rate #
pyear <- sum(elsa_cf$time)
incidencerate <- 1802/pyear
print(pyear)
```

```
## [1] 89155.01
```

```r
print(incidencerate)
```

```
## [1] 0.02021199
```

```r
# This gives a summary of survival time again
summary(Surv(time, death))
```

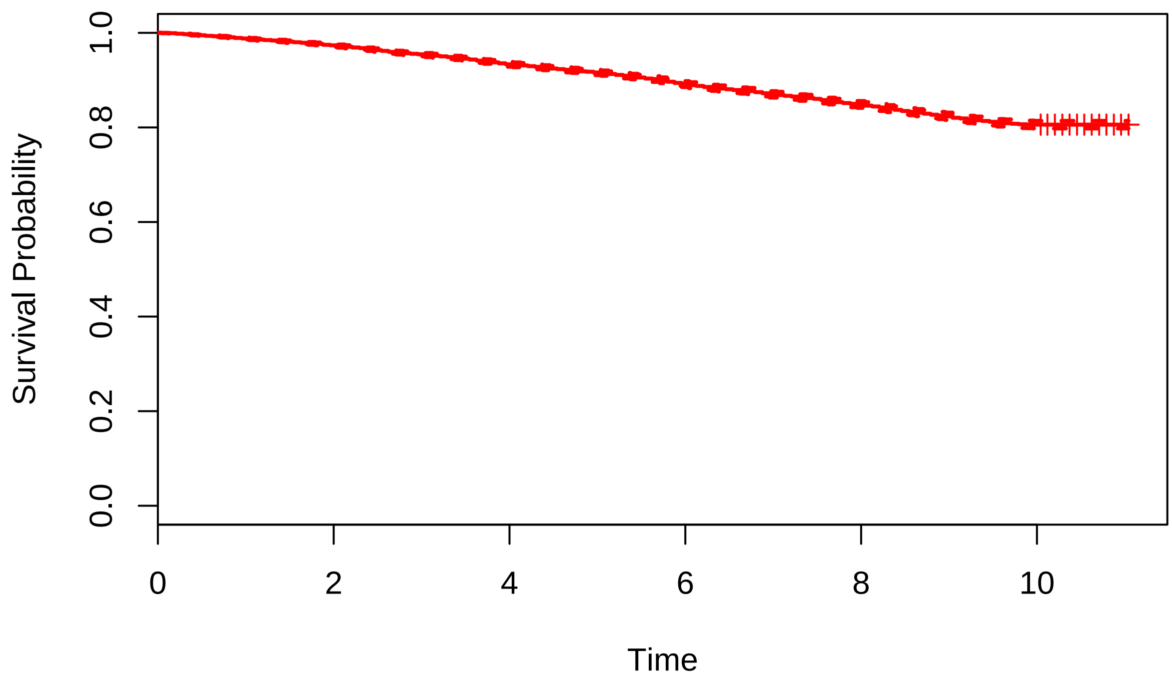
```
##       time              status
##  Min.   : 0.04107   Min.   :0.0000
##  1st Qu.:10.20397   1st Qu.:0.0000
##  Median :10.62286   Median :0.0000
##  Mean   : 9.60515   Mean   :0.1941
##  3rd Qu.:10.78987   3rd Qu.:0.0000
##  Max.   :11.04175   Max.   :1.0000
```

```r
# Create a survival object and return summary #
km <- survfit(Surv(time, death)~ 1)
km
```

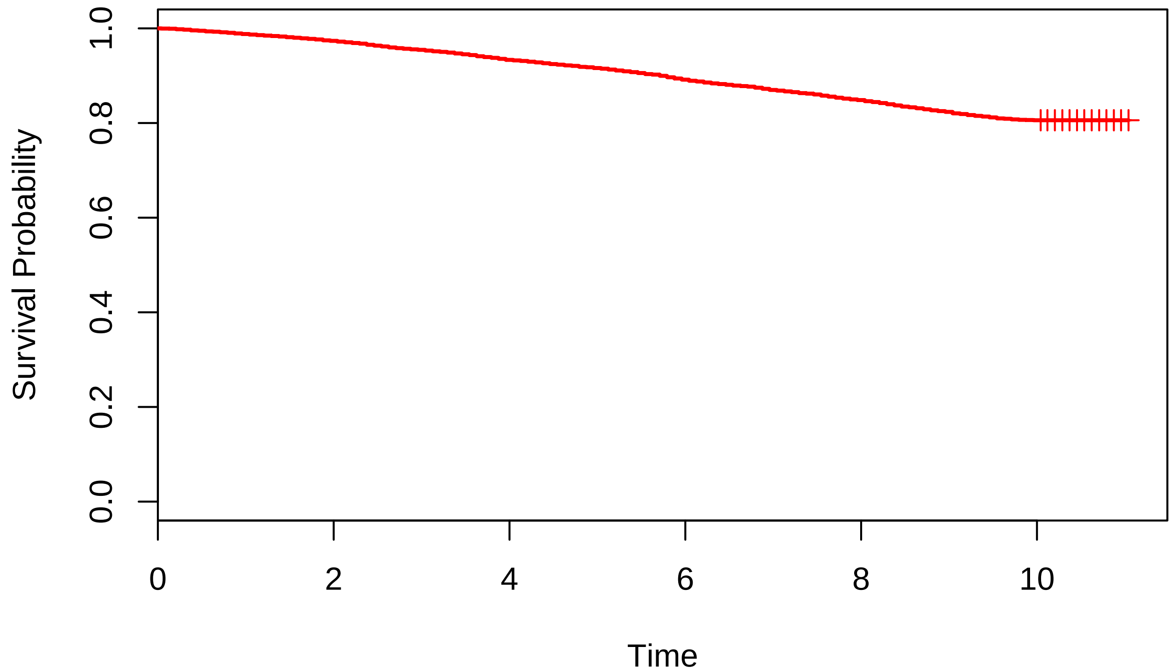
```
## Call: survfit(formula = Surv(time, death) ~ 1)
##
## records   n.max n.start  events  median 0.95LCL 0.95UCL
##    9282    9282    9282    1802      NA      NA      NA
```

```r
# It doesn't give any value for the median survival


# Let's assess this graphically
# Draw the KM plot with CI
plot (km, lty=1, lwd=2, xlab="Time", ylab="Survival Probability",
      col=rainbow(1))
```




```r
# If we don't want Confidence limits
# We have to specify it in survfit
kmnoci <- survfit(Surv(time, death)~ 1, conf.type="none")
plot (kmnoci,lty=1, lwd=2, xlab="Time", ylab="Survival Probability",
      col=rainbow(1))
```




```r
# Quantiles of survival #
quantile(kmnoci)
```

```
## 25 50 75
## NA NA NA
```

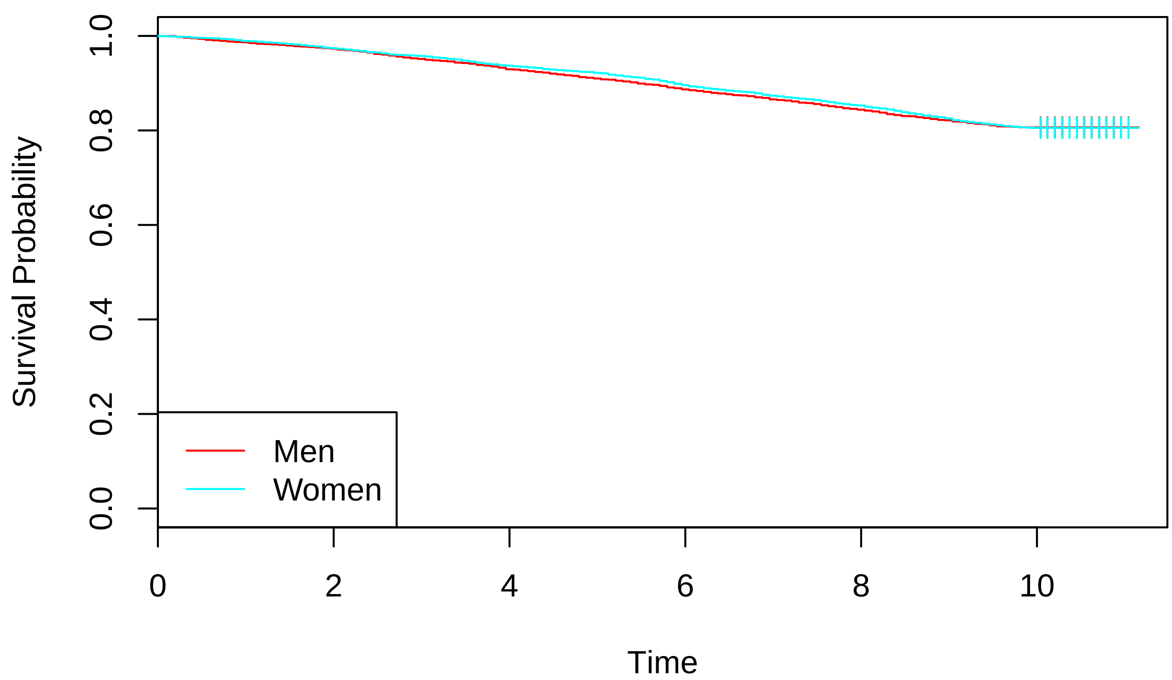
```r
# No value for the 25% either

# We can return the survival table at every time
# or specify at a specific time point, here 11y
# which is the end of follow-up
summary(kmnoci, times=11)
```

```
## Call: survfit(formula = Surv(time, death) ~ 1, conf.type = "none")
##
##  time n.risk n.event survival std.err
##    11     52    1802    0.806 0.00411
```

```r
# We see here that at the end of follow-up
# the survival was ~80% hence less than 25%
# developed the event, which is why R can't estimate
# a 25% survival
#################################################################
## 4. Group comparisons                                         ##
```

```r
##############################################################
# Draw the KM plot by sex
kmsex <- survfit( Surv(time, death)~ strata(sex), data=elsa_cf,
                  conf.type="none")
plot(kmsex, lty=1, lwd=1, xlab="Time", ylab="Survival Probability",
     col=rainbow(2))
legend("bottomleft", c("Men", "Women") , lty=1, lwd=1,
       col=rainbow(2) )
```




```r
#This function implements the G-rho family of Harrington and Fleming (1982),
# with weights on each death of S(t)^rho, where S is the Kaplan-Meier
#estimate of survival.
# With rho = 0 this is the log-rank or Mantel-Haenszel test,
# and with rho = 1 it is equivalent to the Peto & Peto modification
# of the Gehan-Wilcoxon test.
survdiff(Surv(time, death) ~ sex, rho=0)
```

```
## Call:
## survdiff(formula = Surv(time, death) ~ sex, rho = 0)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=0 4484      868      867  0.000575   0.00111
## sex=1 4798      934      935  0.000534   0.00111
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.973
```

```r
# It is not significant: no difference between men and women
##############################################################
## 5. Exercise                                              ##
##############################################################
# Divide cognitive function score (tertiles)
quantile(elsa_cf$cf1, prob=c(0.33, 0.66))
```

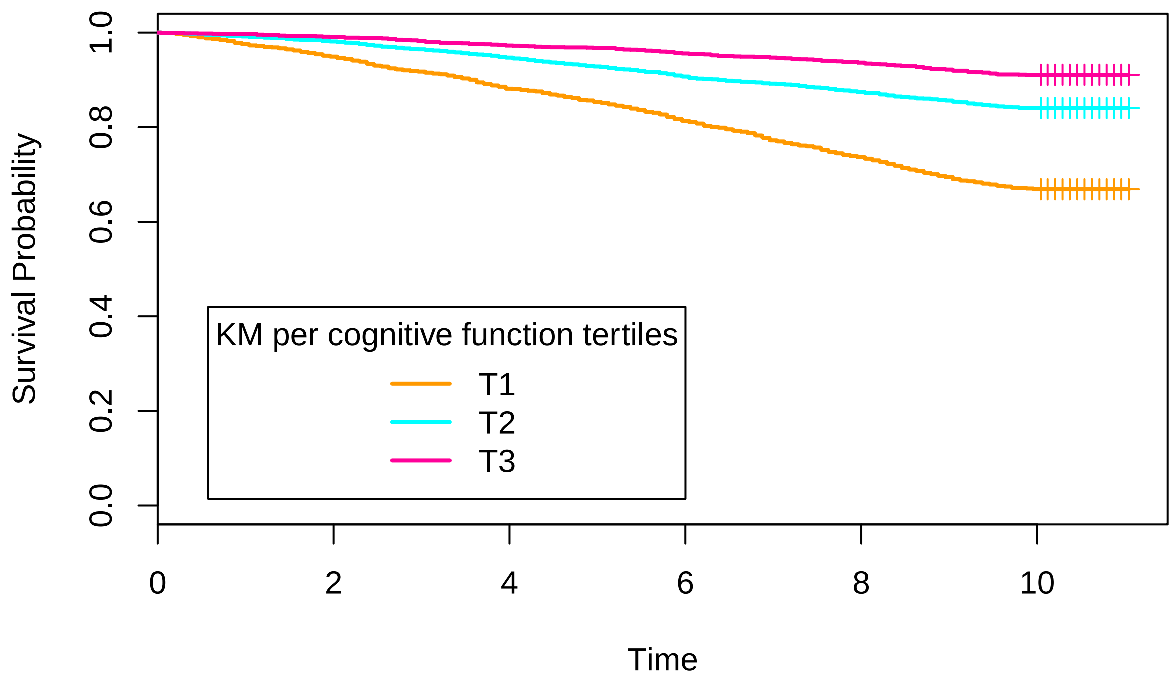
```
## 33% 66%
##  42  52
```

```r
elsa_cf$t_cf1 <-cut(elsa_cf$cf1, breaks=c(0, 42, 52, 194))
# Divide cognitive function score (quintiles)
quantile(elsa_cf$cf1, prob=c(0.20, 0.40, 0.60, 0.80))
```

```
## 20% 40% 60% 80%
##  38  45  50  57
```

```r
elsa_cf$q_cf1 <-cut(elsa_cf$cf1, breaks=c(0, 38, 45, 50, 57, 194))

# Draw the KM plot by tertiles t_cf1
kmcft <- survfit( Surv(time, death)~ strata(t_cf1), data=elsa_cf,
                  conf.type="none")
plot(kmcft, lty=1, lwd=2, xlab="Time", ylab="Survival Probability",
     col=rainbow(3, start=0.1, end=0.9) )
legend("bottomleft", inset=.05, title="KM per cognitive function tertiles",
       c("T1", "T2", "T3"), lty=1, lwd=2 ,col = rainbow(3, start=0.1, end=0.9) )
```

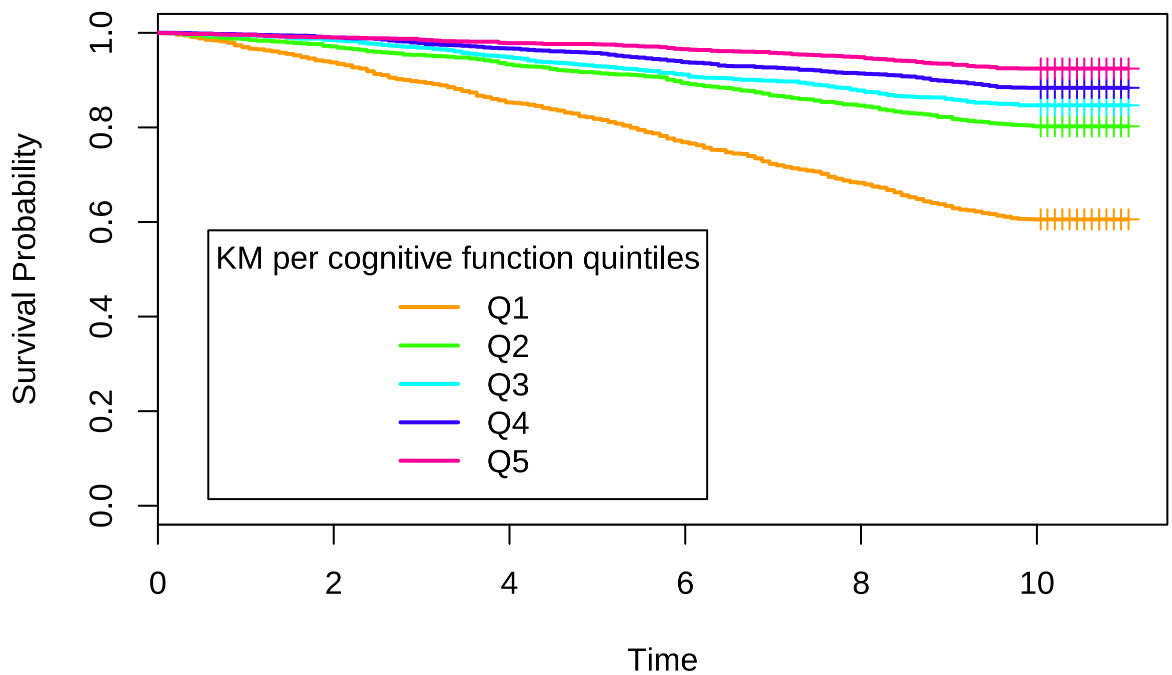



```r
#logrank
survdiff(Surv(elsa_cf$time, elsa_cf$death) ~ elsa_cf$t_cf1, rho=0)
```

```
## Call:
## survdiff(formula = Surv(elsa_cf$time, elsa_cf$death) ~ elsa_cf$t_cf1,
##     rho = 0)
##
##                             N Observed Expected (O-E)^2/E (O-E)^2/V
## elsa_cf$t_cf1=(0,42]    3094     1025      551     407.5     588.4
## elsa_cf$t_cf1=(42,52]   3196      510      632      23.7      36.6
## elsa_cf$t_cf1=(52,194]  2992      267      618     199.7     304.6
##
##  Chisq= 632  on 2 degrees of freedom, p= 0
```

```r
# Significant difference

# Draw the KM plot by quintiles q_cf1
kmcfq <- survfit( Surv(time, death)~ strata(q_cf1), data=elsa_cf,
                  conf.type="none")
plot(kmcfq, lty=1,lwd=2, xlab="Time", ylab="Survival Probability",
     col=rainbow(5,start=0.1, end=0.9) )
legend("bottomleft", inset=.05, title="KM per cognitive function quintiles",
       c("Q1", "Q2", "Q3", "Q4", "Q5"), lty=1, lwd=2,
       col = rainbow(5,start=0.1, end=0.9) )
```




```r
#logrank
survdiff(Surv(elsa_cf$time, elsa_cf$death) ~ elsa_cf$q_cf1, rho=0)
```

```
## Call:
## survdiff(formula = Surv(elsa_cf$time, elsa_cf$death) ~ elsa_cf$q_cf1,
##     rho = 0)
##
##                             N Observed Expected (O-E)^2/E (O-E)^2/V
## elsa_cf$q_cf1=(0,38]    2046      807      348    607.11   754.127
## elsa_cf$q_cf1=(38,45]  1978      391      384      0.14     0.178
## elsa_cf$q_cf1=(45,50]  1632      250      324     16.97    20.712
## elsa_cf$q_cf1=(50,57]  1970      229      402     74.13    95.496
## elsa_cf$q_cf1=(57,194] 1656      125      345    140.31   173.776
##
##  Chisq= 841  on 4 degrees of freedom, p= 0
```

```r
# Significant difference
```