INTRODUCTION-TO-SURVIVAL-ANALYSIS-IN-R-YOUTUBE-VIDEO.R

USER

2024-02-20

library(tidyverse)

## Warning: package 'ggplot2' was built under R version 4.3.2

## Warning: package 'dplyr' was built under R version 4.3.2

## Warning: package 'lubridate' was built under R version 4.3.2

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(survival)

## Warning: package 'survival' was built under R version 4.3.2

library(ggsurvfit)

## Warning: package 'ggsurvfit' was built under R version 4.3.2

library(survminer)

## Warning: package 'survminer' was built under R version 4.3.2

## Loading required package: ggpubr

## Warning: package 'ggpubr' was built under R version 4.3.2

##   
## Attaching package: 'survminer'  
##   
## The following object is masked from 'package:survival':  
##   
## myeloma

library(broom)  
  
aml <- aml  
  
km <- survfit(Surv(time,status) ~ 1,data=aml)  
km

## Call: survfit(formula = Surv(time, status) ~ 1, data = aml)  
##   
## n events median 0.95LCL 0.95UCL  
## [1,] 23 18 27 18 45

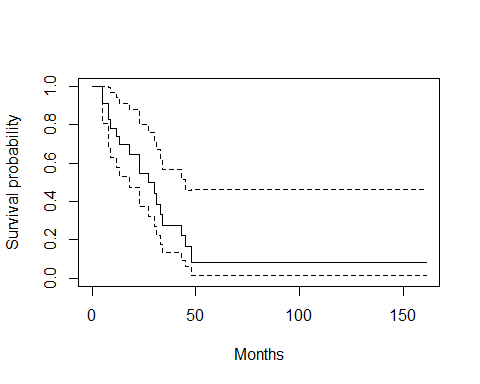
aml |>   
 survfit(Surv(time,status)~1,data=\_)

## Call: survfit(formula = Surv(time, status) ~ 1, data = aml)  
##   
## n events median 0.95LCL 0.95UCL  
## [1,] 23 18 27 18 45

## using tidy() from the broom package to create a tibble  
tidy(km)

## # A tibble: 18 × 8  
## time n.risk n.event n.censor estimate std.error conf.high conf.low  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 5 23 2 0 0.913 0.0643 1 0.805   
## 2 8 21 2 0 0.826 0.0957 0.996 0.685   
## 3 9 19 1 0 0.783 0.110 0.971 0.631   
## 4 12 18 1 0 0.739 0.124 0.942 0.580   
## 5 13 17 1 1 0.696 0.138 0.912 0.531   
## 6 16 15 0 1 0.696 0.138 0.912 0.531   
## 7 18 14 1 0 0.646 0.157 0.878 0.475   
## 8 23 13 2 0 0.547 0.196 0.803 0.372   
## 9 27 11 1 0 0.497 0.218 0.762 0.324   
## 10 28 10 0 1 0.497 0.218 0.762 0.324   
## 11 30 9 1 0 0.442 0.248 0.718 0.272   
## 12 31 8 1 0 0.386 0.282 0.671 0.223   
## 13 33 7 1 0 0.331 0.321 0.622 0.177   
## 14 34 6 1 0 0.276 0.369 0.569 0.134   
## 15 43 5 1 0 0.221 0.432 0.515 0.0947  
## 16 45 4 1 1 0.166 0.519 0.458 0.0598  
## 17 48 2 1 0 0.0828 0.877 0.462 0.0148  
## 18 161 1 0 1 0.0828 0.877 0.462 0.0148

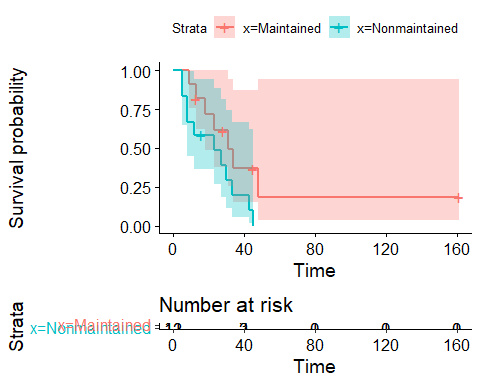
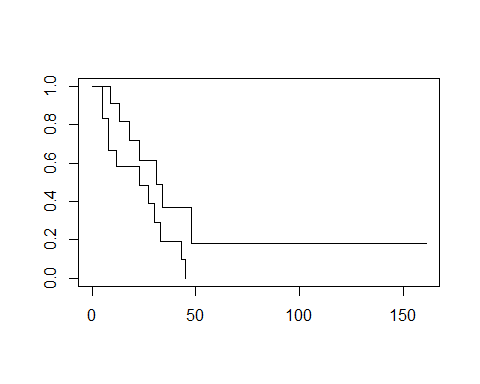
km\_table <- tidy(km)  
  
plot(km,xlab="Months",ylab="Survival probability")



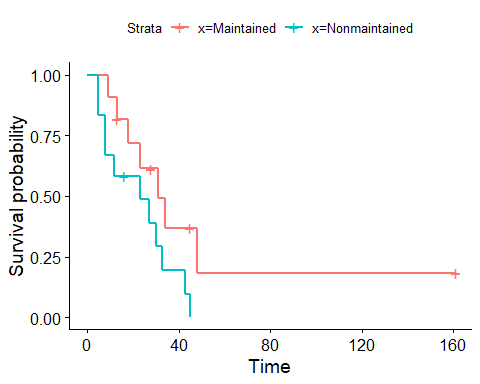
km2 <- aml |>   
 survfit(Surv(time,status)~x,data=\_)  
  
km\_table2 <- tidy(km2)  
km\_table2

## # A tibble: 20 × 9  
## time n.risk n.event n.censor estimate std.error conf.high conf.low strata   
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 9 11 1 0 0.909 0.0953 1 0.754 x=Mainta…  
## 2 13 10 1 1 0.818 0.142 1 0.619 x=Mainta…  
## 3 18 8 1 0 0.716 0.195 1 0.488 x=Mainta…  
## 4 23 7 1 0 0.614 0.249 0.999 0.377 x=Mainta…  
## 5 28 6 0 1 0.614 0.249 0.999 0.377 x=Mainta…  
## 6 31 5 1 0 0.491 0.334 0.946 0.255 x=Mainta…  
## 7 34 4 1 0 0.368 0.442 0.875 0.155 x=Mainta…  
## 8 45 3 0 1 0.368 0.442 0.875 0.155 x=Mainta…  
## 9 48 2 1 0 0.184 0.834 0.944 0.0359 x=Mainta…  
## 10 161 1 0 1 0.184 0.834 0.944 0.0359 x=Mainta…  
## 11 5 12 2 0 0.833 0.129 1 0.647 x=Nonmai…  
## 12 8 10 2 0 0.667 0.204 0.995 0.447 x=Nonmai…  
## 13 12 8 1 0 0.583 0.244 0.941 0.362 x=Nonmai…  
## 14 16 7 0 1 0.583 0.244 0.941 0.362 x=Nonmai…  
## 15 23 6 1 0 0.486 0.305 0.883 0.268 x=Nonmai…  
## 16 27 5 1 0 0.389 0.378 0.816 0.185 x=Nonmai…  
## 17 30 4 1 0 0.292 0.476 0.741 0.115 x=Nonmai…  
## 18 33 3 1 0 0.194 0.627 0.664 0.0569 x=Nonmai…  
## 19 43 2 1 0 0.0972 0.945 0.620 0.0153 x=Nonmai…  
## 20 45 1 1 0 0 Inf NA NA x=Nonmai…

plot(km2)  
  
  
## using ggsurvplot from survminer  
ggsurvplot(km2,conf.int = TRUE,  
 risk.table = TRUE)



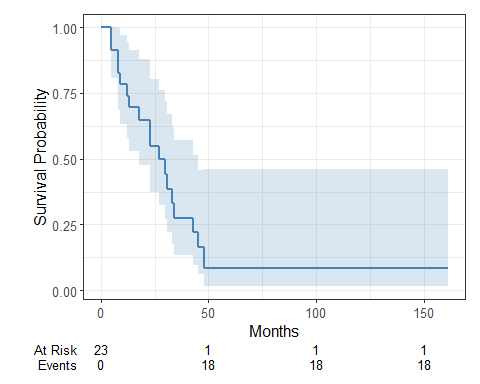
ggsurvplot(km2)



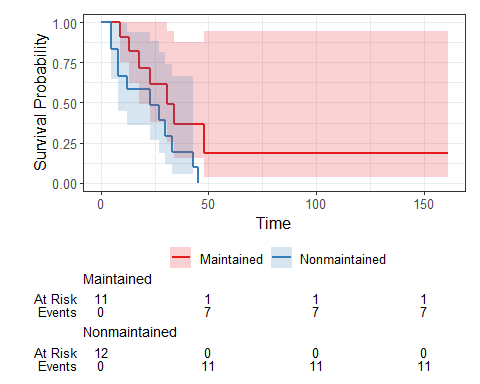
## to calculate the Logrank statistic to compare the two curves  
## Ho : survival across the two groups are equivalent  
  
survdiff(Surv(time,status)~x,data=aml)

## Call:  
## survdiff(formula = Surv(time, status) ~ x, data = aml)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## x=Maintained 11 7 10.69 1.27 3.4  
## x=Nonmaintained 12 11 7.31 1.86 3.4  
##   
## Chisq= 3.4 on 1 degrees of freedom, p= 0.07

##  
  
  
  
  
  
  
  
##using ggsurvfit  
km |>   
 ggsurvfit(color='steelblue',linewidth=1)+  
 add\_risktable()+  
 add\_confidence\_interval(fill="steelblue")+  
 labs(x="Months")



aml |>   
survfit2(Surv(time,status)~x,data=\_) |>   
 ggsurvfit(linewidth=1)+  
 add\_risktable()+  
 scale\_color\_brewer(palette="Set1")+  
 add\_confidence\_interval()+  
 scale\_fill\_brewer(palette = "Set1")



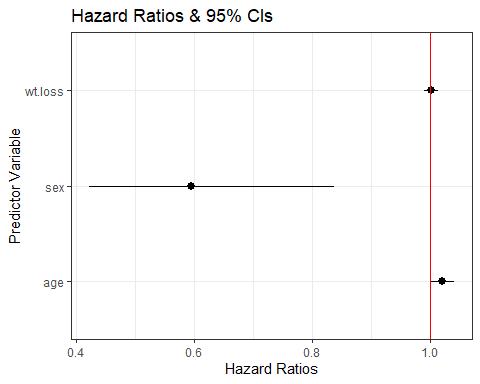
## if you want to fit KM for the whole dataset you use ~1   
## if you want to check survival by treatment groups,use ~treatment  
  
  
  
  
# COX PROPORTIONAL HAZARDS MODEL ------------------------------------------  
  
lung <- lung  
## t =survival time in days  
## status 1= censored, 2 = dead  
## sex 1=male, 2 = female  
## wt.loss = weight loss in pounds in last 6 months  
#3 age = age in years ( assessed at beginning)  
  
## Cox model = coxph(formula,data= )  
## formula = Surv(time, event) ~x1 + x2 + ..xp  
  
lung\_cox <- coxph(Surv(time,status) ~age + sex + wt.loss,data = lung)  
summary(lung\_cox)

## Call:  
## coxph(formula = Surv(time, status) ~ age + sex + wt.loss, data = lung)  
##   
## n= 214, number of events= 152   
## (14 observations deleted due to missingness)  
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## age 0.0200882 1.0202913 0.0096644 2.079 0.0377 \*   
## sex -0.5210319 0.5939074 0.1743541 -2.988 0.0028 \*\*  
## wt.loss 0.0007596 1.0007599 0.0061934 0.123 0.9024   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## age 1.0203 0.9801 1.0011 1.0398  
## sex 0.5939 1.6838 0.4220 0.8359  
## wt.loss 1.0008 0.9992 0.9887 1.0130  
##   
## Concordance= 0.612 (se = 0.027 )  
## Likelihood ratio test= 14.67 on 3 df, p=0.002  
## Wald test = 13.98 on 3 df, p=0.003  
## Score (logrank) test = 14.24 on 3 df, p=0.003

## tidying up cox regression output  
  
lungcoxtab <- tidy(lung\_cox,exponentiate=TRUE,conf.int = TRUE)  
lungcoxtab

## # A tibble: 3 × 7  
## term estimate std.error statistic p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 age 1.02 0.00966 2.08 0.0377 1.00 1.04   
## 2 sex 0.594 0.174 -2.99 0.00280 0.422 0.836  
## 3 wt.loss 1.00 0.00619 0.123 0.902 0.989 1.01

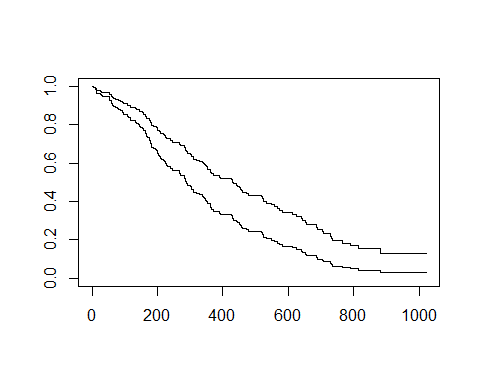
## Plotting the Hazard ratios  
lungcoxtab |>   
 ggplot(aes(y=term,x=estimate,xmin=conf.low,xmax=conf.high))+  
 geom\_pointrange()+  
 geom\_vline(xintercept=1,color="red")+  
 labs(x="Hazard Ratios",title="Hazard Ratios & 95% CIs",y= "Predictor Variable")+  
 theme\_bw()



## Predicting survival after coxph() with survfit()  
  
## predict survival fxn for subject with mean values on all covariates  
## However predicting survival at a mean o all covariates may not make sense  
## especialy if one or more of the covariates are factors(categorical)  
## instead it is recommended to supply to always supply a data frame of covariate values   
## at which to predict the survival function to the newdata =() option of survfit()  
  
## making a new data frame  
plotdata <- data.frame(age=mean(lung$age,na.rm=TRUE),  
 sex=1:2,  
 wt.loss=mean(lung$wt.loss,na.rm=TRUE))  
  
plotdata

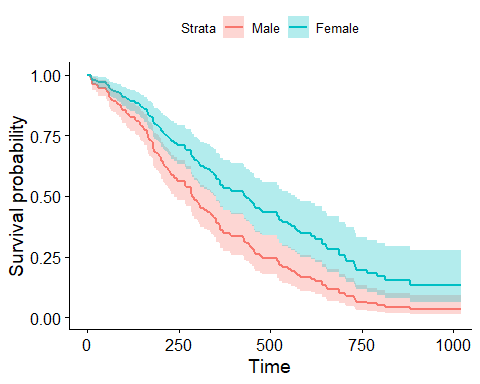
## age sex wt.loss  
## 1 62.44737 1 9.831776  
## 2 62.44737 2 9.831776

## supplying newdata frame to survfit  
surv\_by\_sex <- survfit(lung\_cox,newdata = plotdata)  
  
survbysex\_tab <- tidy(surv\_by\_sex)  
  
plot(surv\_by\_sex)



ggsurvplot(surv\_by\_sex,data=plotdata,censor=FALSE,legend.labs=c("Male","Female"))

## Warning: `gather\_()` was deprecated in tidyr 1.2.0.  
## ℹ Please use `gather()` instead.  
## ℹ The deprecated feature was likely used in the survminer package.  
## Please report the issue at <https://github.com/kassambara/survminer/issues>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



## Assesing the proportional hazards assumptions  
  
## 1) Chi-squared test based on Scheonfeld residuals is available with cox.zph()  
##to test the hypothesis Ho:covariate efect is constant(proportional) over time  
## Ha: covariate effect changes over time  
  
cox.zph(lung\_cox)

## chisq df p  
## age 0.5077 1 0.48  
## sex 2.5489 1 0.11  
## wt.loss 0.0144 1 0.90  
## GLOBAL 3.0051 3 0.39

plot(cox.zph(lung\_cox),col="steelblue")

