

# R code for Survival Lecture 3

## Contents

Plot KM curve and the predicted survival probability under Cox model . . . . .	2
Generate log-minus-log plot . . . . .	3
Check the proportional hazards assumption . . . . .	4
Stratified Cox regression . . . . .	6
Time-dep cov approach for dealing with violation of PH assumption . . . . .	7
Model checking with martingale residuals . . . . .	8

## Plot KM curve and the predicted survival probability under Cox model

- Load all the R packages

```
packages_to_load <- c("haven", "survival", "survminer", "pander")  
lapply(packages_to_load, library, character.only = TRUE)
```

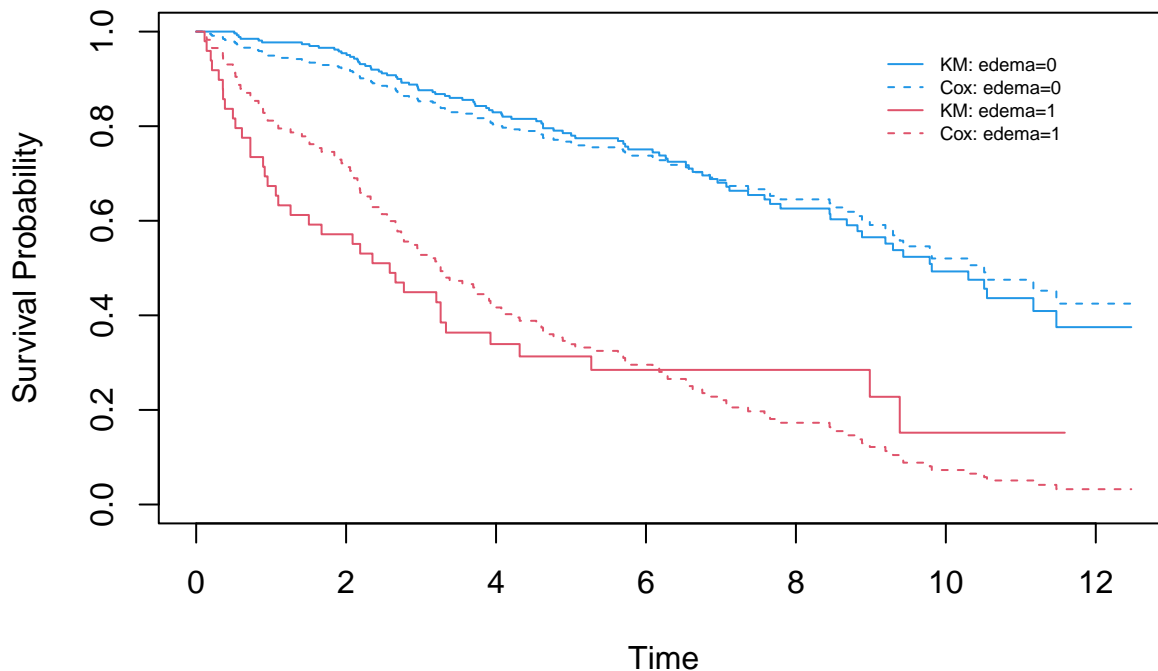
- Load the PMC dataset

```
PBC <- read_dta("pbc.dta")
```

- Generate a plot that displays the Kaplan-Meier survival curves and predicted survival under the Cox model.

```
km.fit <- survfit(Surv(years, status) ~ edema, data=PBC)  
cox.fit <- coxph(Surv(years, status) ~ edema, data=PBC)  
newdata <- data.frame(edema=c(0,1))  
predict.fit <- survfit(cox.fit, newdata= newdata)  
  
plot(km.fit, col=c(4,2), xlab = "Time", ylab = "Survival Probability",  
     main = "Kaplan-Meier and Cox Fit by Edema")  
legend(9, 0.98, c("KM: edema=0", "Cox: edema=0", "KM: edema=1", "Cox: edema=1"),  
     col=c(4,4,2,2), lty=c(1,2,1,2), bty="n", cex=0.6)  
lines(predict.fit, col=c(4,2), lty=c(2,2))
```

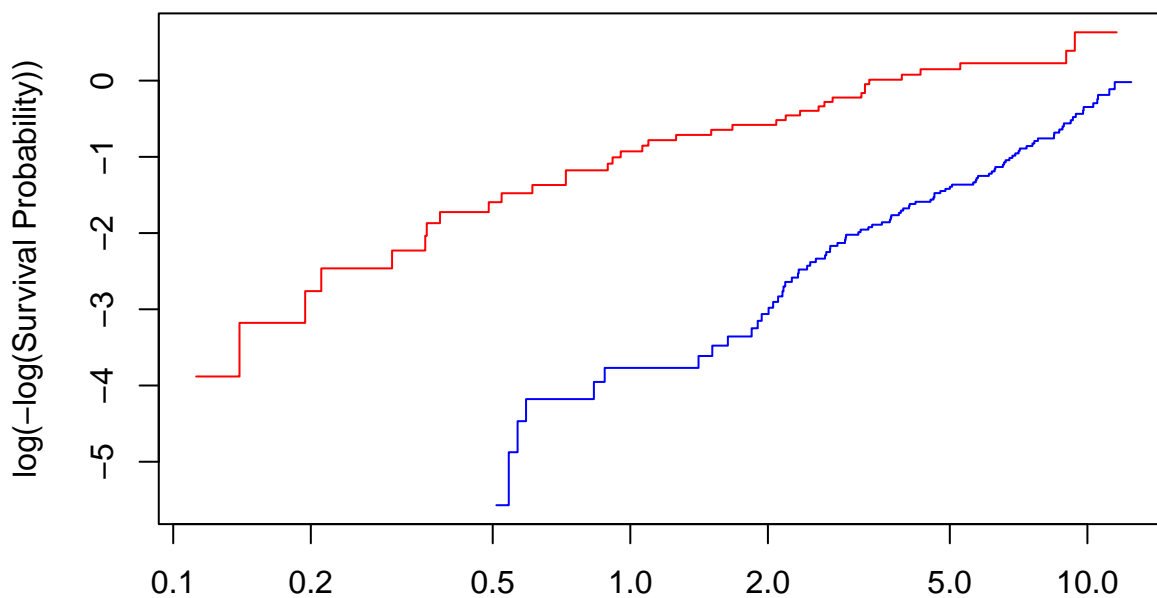
### Kaplan-Meier and Cox Fit by Edema



## Generate log-minus-log plot

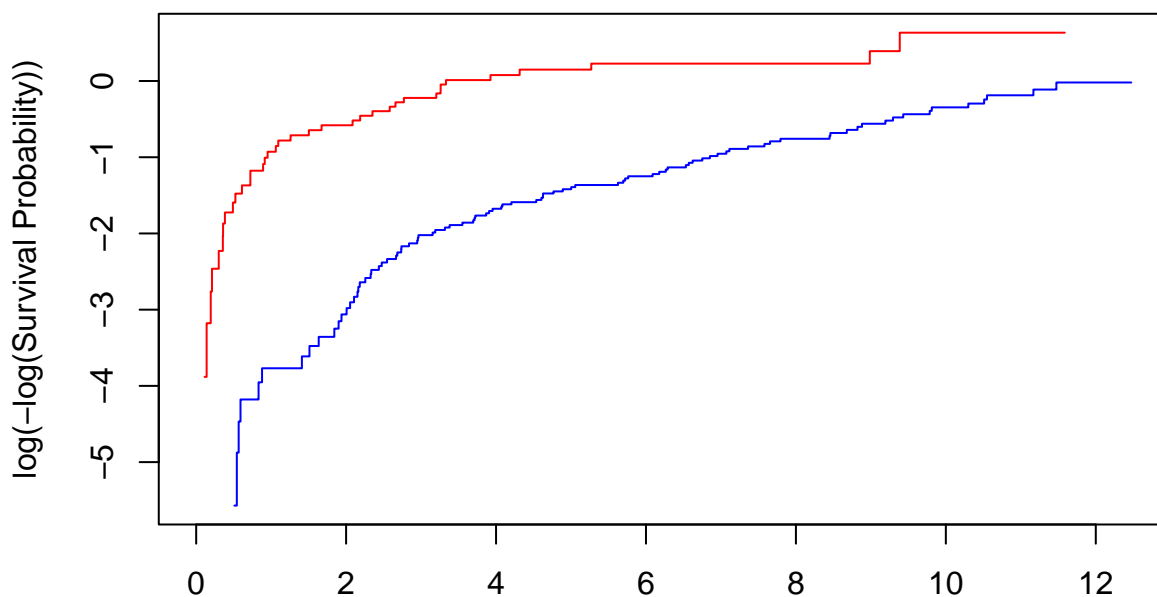
- Create a log-minus-log plot with a logarithmic scale on the x-axis.

```
plot(survfit(Surv(years, status) ~ edema, data=PBC),  
     ylab="log(-log(Survival Probability))",  
     col=c("blue", "red"), fun="cloglog", log="")
```



- Create a log-minus-log plot without a logarithmic scale on the x-axis.

```
myfun=function(p){return(log(-log(p)))}  
plot(survfit(Surv(years, status) ~ edema, data=PBC),  
     ylab="log(-log(Survival Probability))",  
     col=c("blue", "red"), fun=myfun)
```



## Check the proportional hazards assumption

- Schoenfeld test for testing the proportional hazards assumption

The table component provides the results of a formal score test for slope=0, a linear fit to the plot would approximate the test.

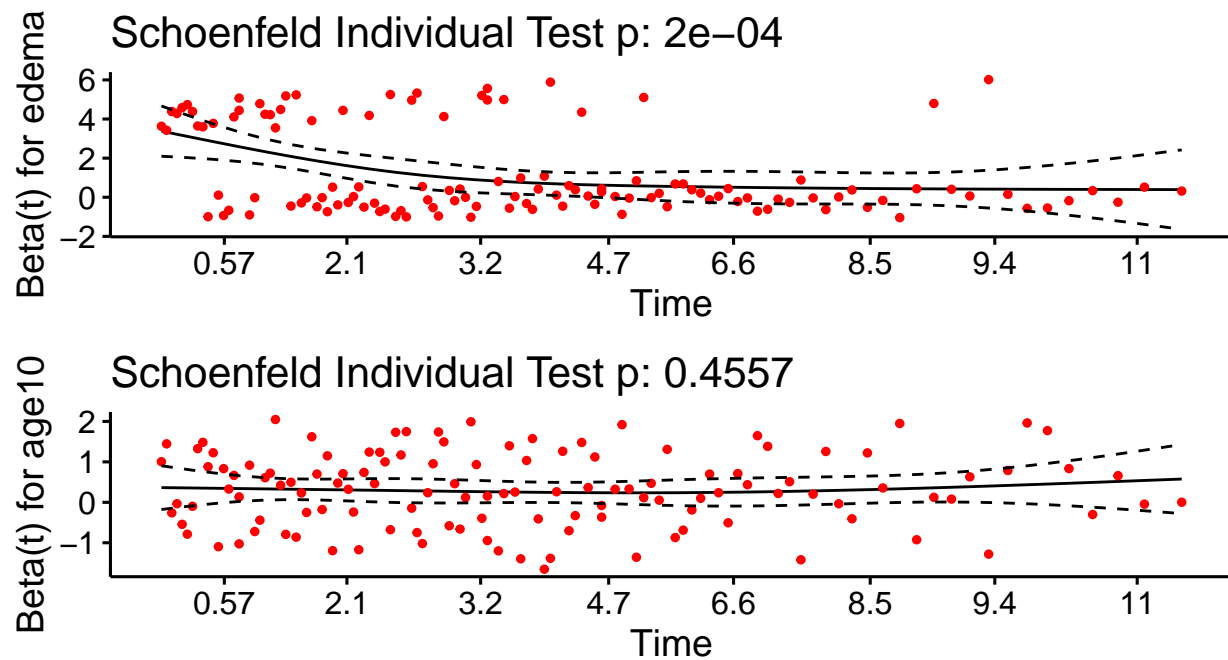
```
PBC$age10 <- PBC$age/10
cox.fit1 <- coxph(Surv(years, status) ~ edema + age10, data=PBC)
# Schoenfeld test
test.ph <- cox.zph(cox.fit1)
test.ph
```

```
##           chisq df          p
## edema  13.842  1 0.00020
## age10   0.556  1 0.45569
## GLOBAL 13.843  2 0.00099
```

- Smooth scaled Schoenfeld residuals vs. time. The plot gives an estimate of the time-dependent coefficient. If the proportional hazards assumption holds then the true function would be a horizontal line.

```
ggcoxzph(test.ph)
```

Global Schoenfeld Test p: 0.0009862

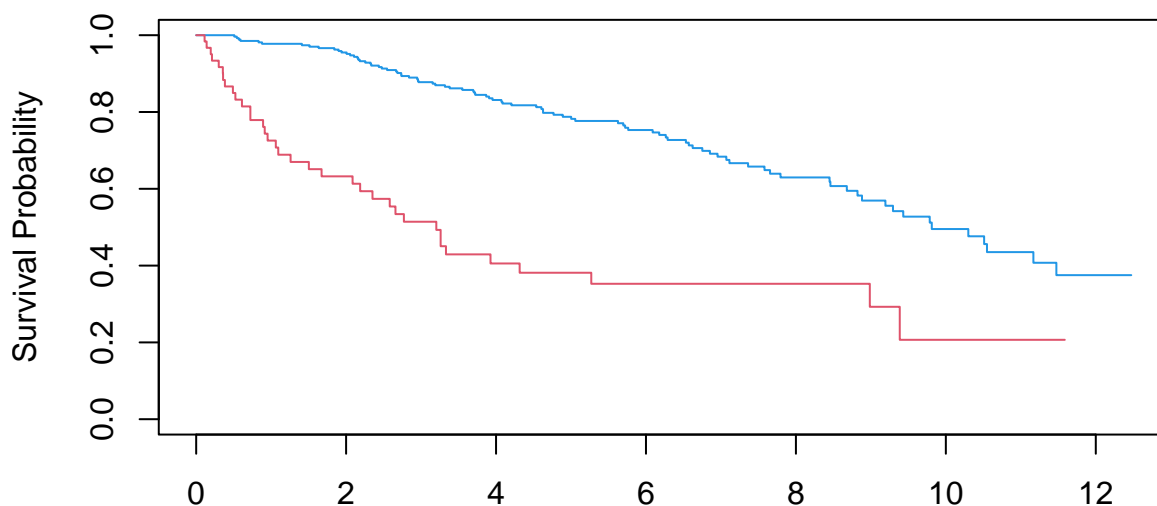


```
#plot(test.ph)
```

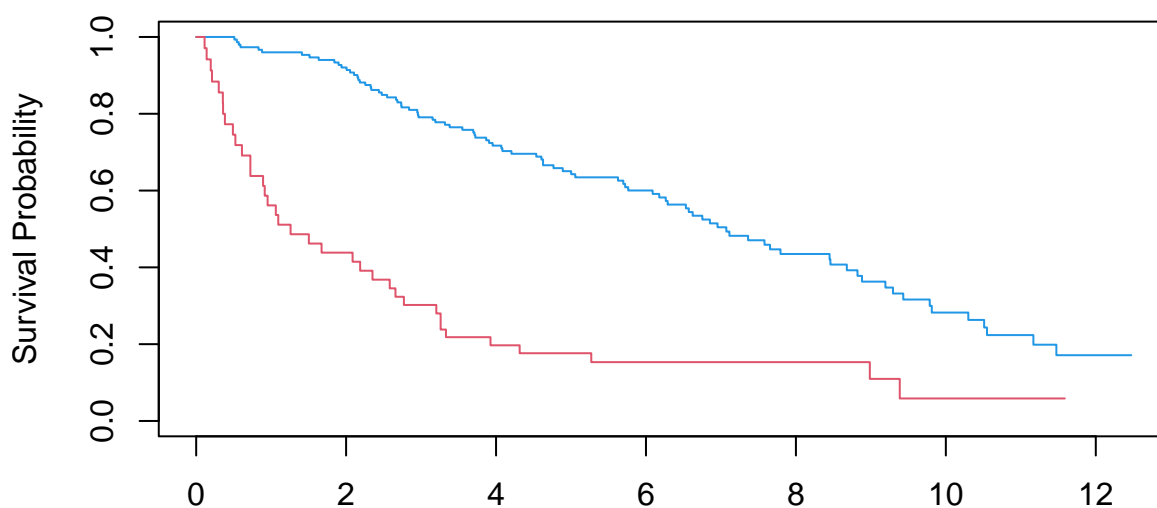
## Stratified Cox regression

Plot the predicted survival function under the stratified Cox model in each edema group by fixing age at 50 and 70.

```
cox.strata <- coxph(Surv(years, status) ~ age10 + strata(edema), data = PBC)
newdata1 <- data.frame(edema=c(0,1), age10=5)
predict.strata1 <- survfit(cox.strata, newdata= newdata1)
plot(predict.strata1, col=c(4,2), ylab="Survival Probability")
```



```
newdata2 <- data.frame(edema=c(0,1), age10=7)
predict.strata2 <- survfit(cox.strata, newdata= newdata2)
plot(predict.strata2, col=c(4,2), ylab="Survival Probability")
```



## Time-dep cov approach for dealing with violation of PH assumption

The `survSplit` function is used to split the data into subintervals based on the years variable. The variables “tstart” and “years” indicate the lower and upper limits of the subintervals, and “grp” indicate the order of subintervals.

```
PBC.td <- survSplit(Surv(years, status) ~., data=PBC,  
                   cut=c(1, 3, 5), episode ="grp")
```

This set of commands generates 4 separate edema variables specific to each time interval;

```
PBC.td$edema01 <- PBC.td$edema * (PBC.td$grp==1)  
PBC.td$edema13 <- PBC.td$edema * (PBC.td$grp==2)  
PBC.td$edema35 <- PBC.td$edema * (PBC.td$grp==3)  
PBC.td$edema5p <- PBC.td$edema * (PBC.td$grp==4)
```

`Surv(tstart, years, status)` specifies the time-to-event data for each patient, where `tstart` is the starting time of the subinterval, `years` is the end time of subinterval, and `status` is an indicator variable for the event of interest at the end of the subinterval.

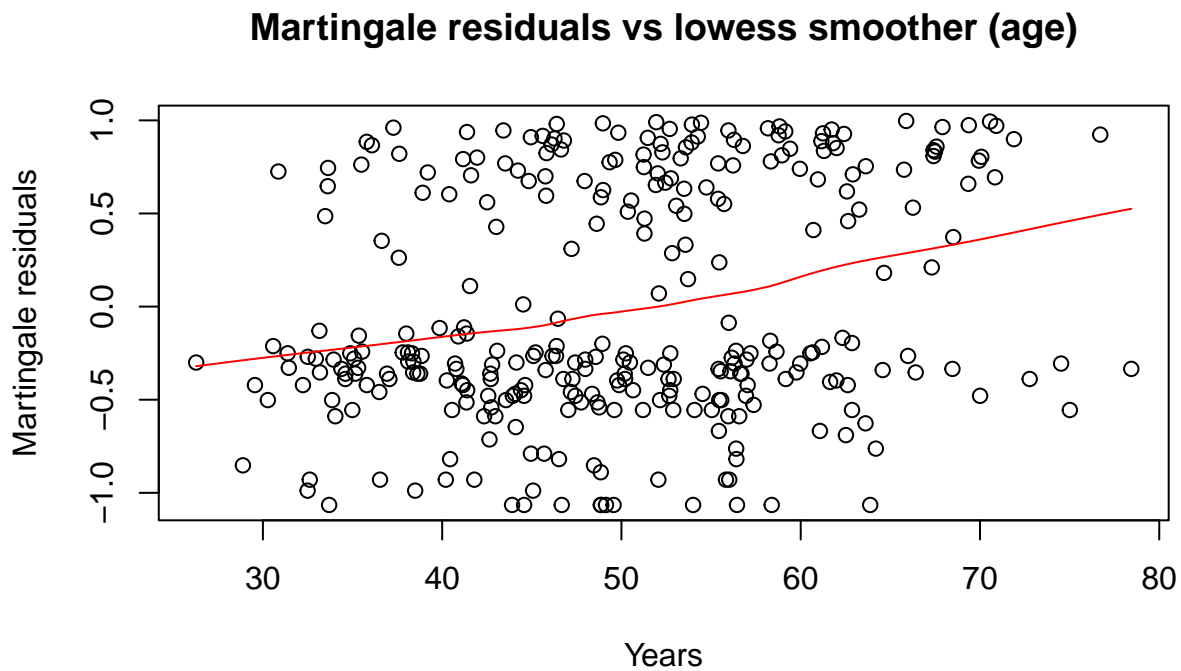
```
cox.timedep <- coxph(Surv(tstart, years, status) ~  
                    edema01 + edema13 + edema35 + edema5p +  
                    age10, data = PBC.td)
```

## Model checking with martingale residuals

The resulting `cox.null` object is a Cox proportional hazards model object that can be used to calculate martingale residuals. The formula “`~1`” indicates that there are no covariates in the model.

```
cox.null <- coxph(Surv(years, status) ~1, data = PBC)

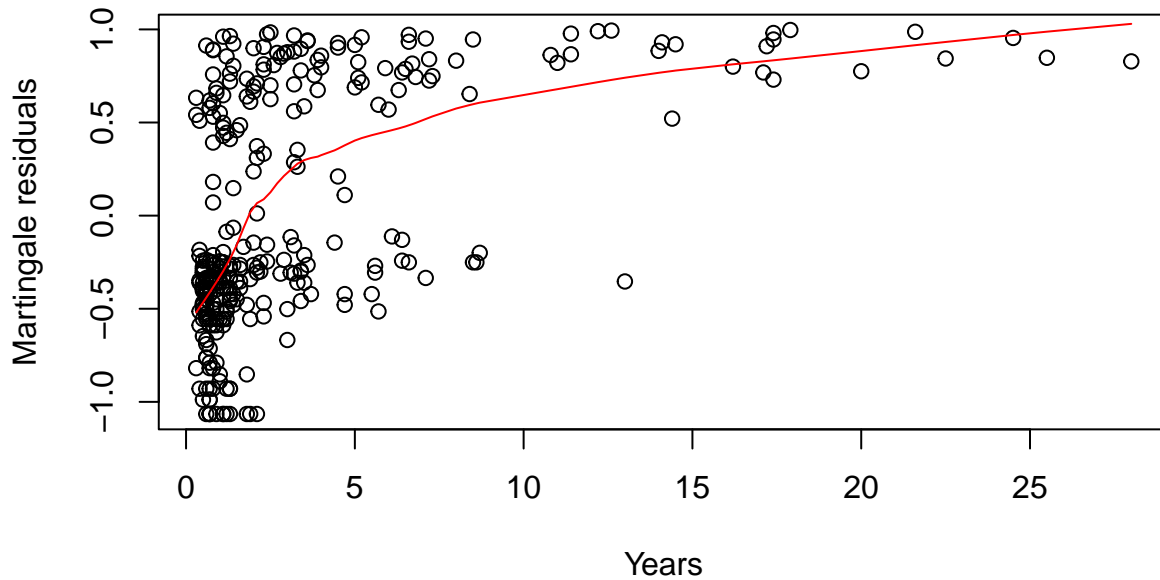
plot(PBC$age, resid(cox.null), xlab="Years",
     ylab="Martingale residuals",
     main="Martingale residuals vs lowess smoother (age)")
lines(lowess(PBC$age, resid(cox.null)), col='red')
```





```
plot(PBC$bilirubin, resid(cox.null), xlab="Years",
     ylab="Martingale residuals",
     main="Martingale residuals vs lowess smoother (bilirubin)")
lines(lowess(PBC$bilirubin, resid(cox.null)),col='red')
```

### Martingale residuals vs lowess smoother (bilirubin)



```
plot(PBC$logbili, resid(cox.null), xlab="Years",
     ylab="Martingale residuals",
     main="Martingale residuals vs lowess smoother (logbili)")
lines(lowess(PBC$logbili, resid(cox.null)),col='red')
```

### Martingale residuals vs lowess smoother (logbili)

