**ADA Homework 11**

**Wenxin Liang wl2455**

Consider the Mayo Clinic Lung Cancer Data in R package survival: data (lung) or data (cancer): including the variables

inst: Institution code

time: Survival time in days

status: censoring status 1=censored, 2=dead

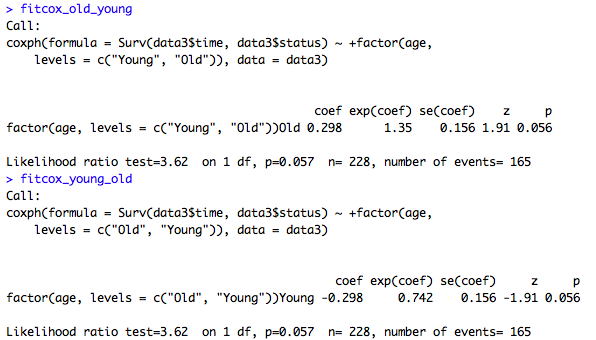
age: Age in years

sex: Male=1 Female=2, etc.

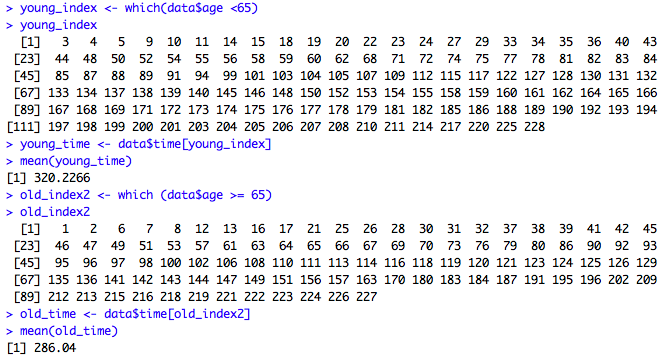
Define ‘AGE GROUP’ as ‘YOUNG” if ‘age < 65”, and OLD, otherwise.

1. Using a Cox proportional hazards model, estimate the hazard rate for old relative to young

As R code showed,



Also based on R, for old relative to young, which may be old over young.



The mean of time for age young is greater than the mean of time for age old then we conclude that the hazard ratio for age old relative to age young is 0.742.

2. Assess the validity of the proportional hazards assumption in (1)

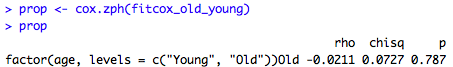
For the proportional hazards assumption we test whether the ratio of hazards is independent of t. Suppose is an indicator variable, taking the values 0 or 1.

Defining the auxiliary variable may test this,

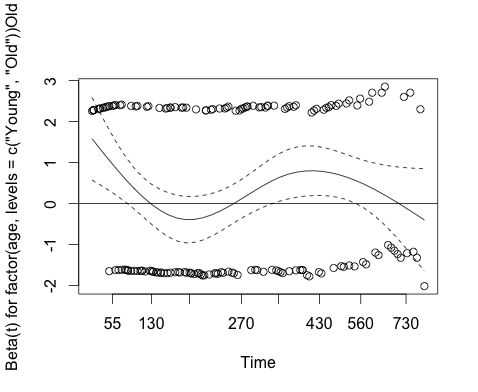
Then the hazard ratio becomes,

Thus, the hypotheses are followed,

Based on R, the result and the plot showed,



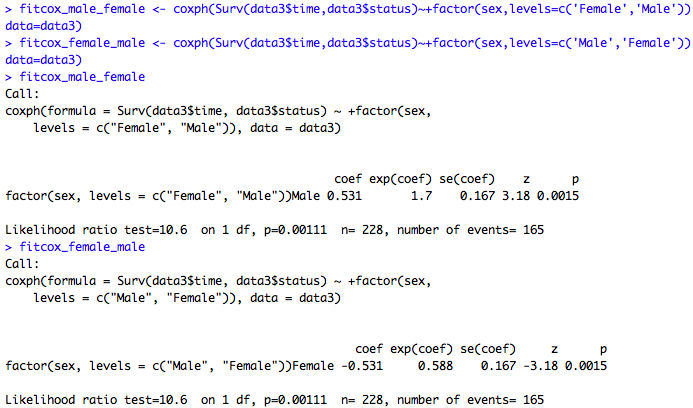
Since p-value >0.05 we fail to reject the null hypothesis then we conclude that the proportional hazards assumption is true. There is evidence of proportional hazard for old relative to young.



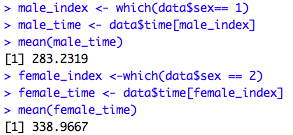
The plot showed the scaled Schoenfeld residuals against transformed time for the covariate, young, in the model fit to the data. The solid line is a smoothing-spline fit to the plot, with the broken lines representing a ± 2-standard-error band around the fit. Since the plot is not systematic departures from a horizontal line, then we can conclude that the proportional hazard assumption is true.

3. Repeat 1, adjusting “Sex”.

Based on R,



Also we calculate the mean of the time for female and male, as followed,



The mean of time for female is greater than the mean of time for male then we conclude that the hazard ratio for age old relative to age young is 0.588.

Following code,

#ADA Homework 11

library(survival)

data = lung

# Question 1

data3 <- data

data3$age[which(data3$age<65)] <- 1

data3$age[which(data3$age>65)] <- 0

data3$age[which(data3$age==65)] <- 0

data3$age[which(data3$age == 0)] <- "Old"

data3$age[which(data3$age == "1")] <- "Young"

fitcox\_old\_young <- coxph(Surv(data3$time,data3$status)~+factor(age,levels=c('Young','Old')),data=data3)

fitcox\_young\_old <- coxph(Surv(data3$time,data3$status)~+factor(age,levels=c('Old','Young')),data=data3)

young\_index <- which(data$age <65)

young\_index

young\_time <- data$time[young\_index]

mean(young\_time)

old\_index2 <- which (data$age >= 65)

old\_index2

old\_time <- data$time[old\_index2]

mean(old\_time)

# young >old

# choose smaller one

# old/young

# 0.7419

# Question 2

prop <- cox.zph(fitcox\_old\_young)

plot(prop)

score <- resid(fitcox\_old\_young,type="score")

score\_frame <- data.frame(score)

plot(data$age,score\_frame[,1],ylab="Score Residuals",xlab ="AGE")

# Question 3

data3$sex[which(data3$sex == 1)] <- "Male"

data3$sex[which(data3$sex == "2")] <- "Female"

fitcox\_male\_female <- coxph(Surv(data3$time,data3$status)~+factor(sex,levels=c('Female','Male')),data=data3)

fitcox\_female\_male <- coxph(Surv(data3$time,data3$status)~+factor(sex,levels=c('Male','Female')),data=data3)

fitcox\_male\_female1 <- coxph(Surv(data$time,data$status)~sex,data=data)

summary(fitcox\_male\_female1)

male\_index <- which(data$sex== 1)

male\_time <- data$time[male\_index]

mean(male\_time)

female\_index <-which(data$sex == 2)

female\_time <- data$time[female\_index]

mean(female\_time)