P8108 Homework 8

Ryan Wei, rw2844

2022-11-17

Problem 1

Here I did some data manipulation to get the wait time for each patient. I noticed that the age was centered by 48, therefore I recovered it to it original scale. For age at transplant, for those who received transplantation, I use the original scale of the baseline age plus the wait time divided by 365.25, that is, I assume the wait time was in days; For those who didn't received transplantation, the age at transplant is the same to the baseline age.

I include the indicator of transplantation and the time-varying variable `AGE_T`, the age at transplant as covariates:

Table 1: Regresion Coefficients Estimates of the Time-Varying Cox Model

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
Transplant Status = 1	-1.7947	0.1662	0.2717	-6.6048	0e+00
Age(t)	0.0599	1.0617	0.0153	3.9223	1e-04

From the regression coefficients result above, we can see that when controlling the age at transplant, the transplant's effect shows significant, and it significantly lower the risk of event, the risk reduction is about 83.4%.
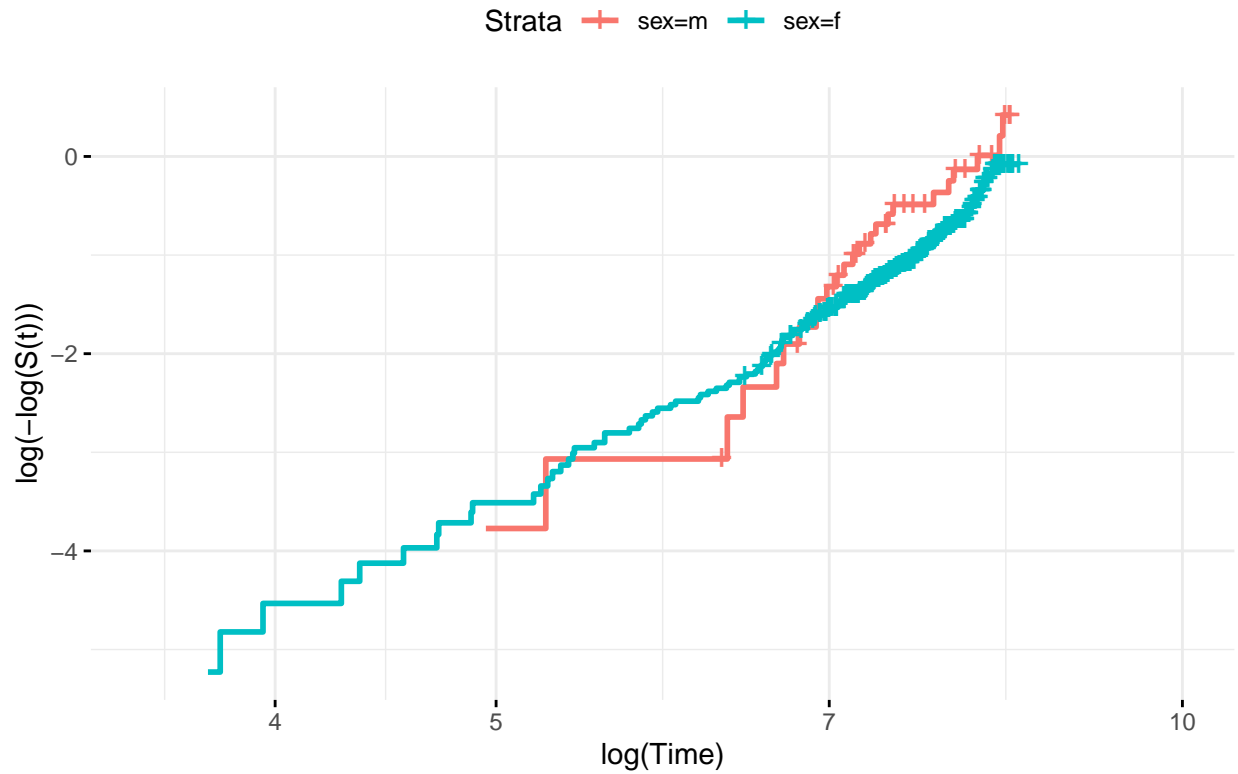
Problem 2

The result of the checking of proportional hazard assumption using $\log\{-\log S(t, Z)\}$ is shown in the following plot:

Product Limit Survival Estimates
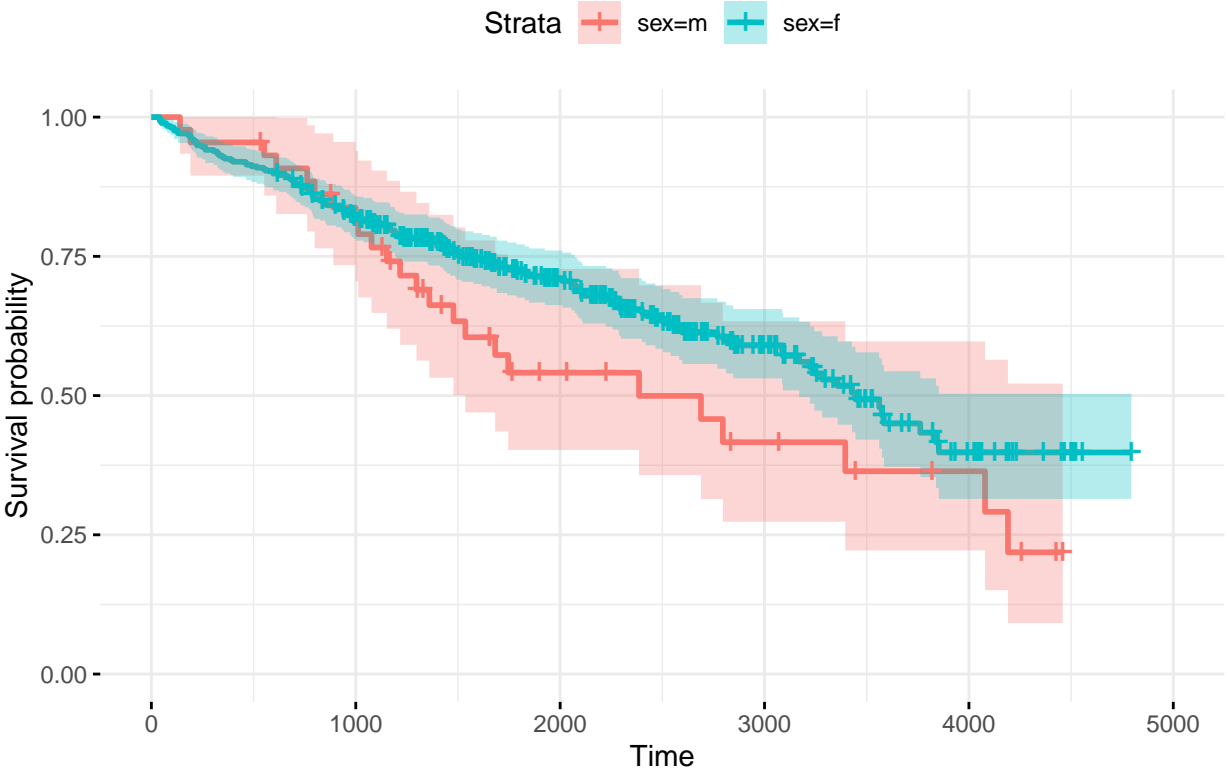
Log of Negative Log of Estimated Survival Function

From the plot above, we can found that the $\log\{-\log S(t, Z)\}$ are not two parallel lines, there exists a cross-over at about time $= 1000$. That means the hazard ratio between male and female is not proportional.

The result of the checking of proportional hazard assumption using the observed and fitted survival function is shown in the following plot:

Observed vs. Fitted Survival, By Sex
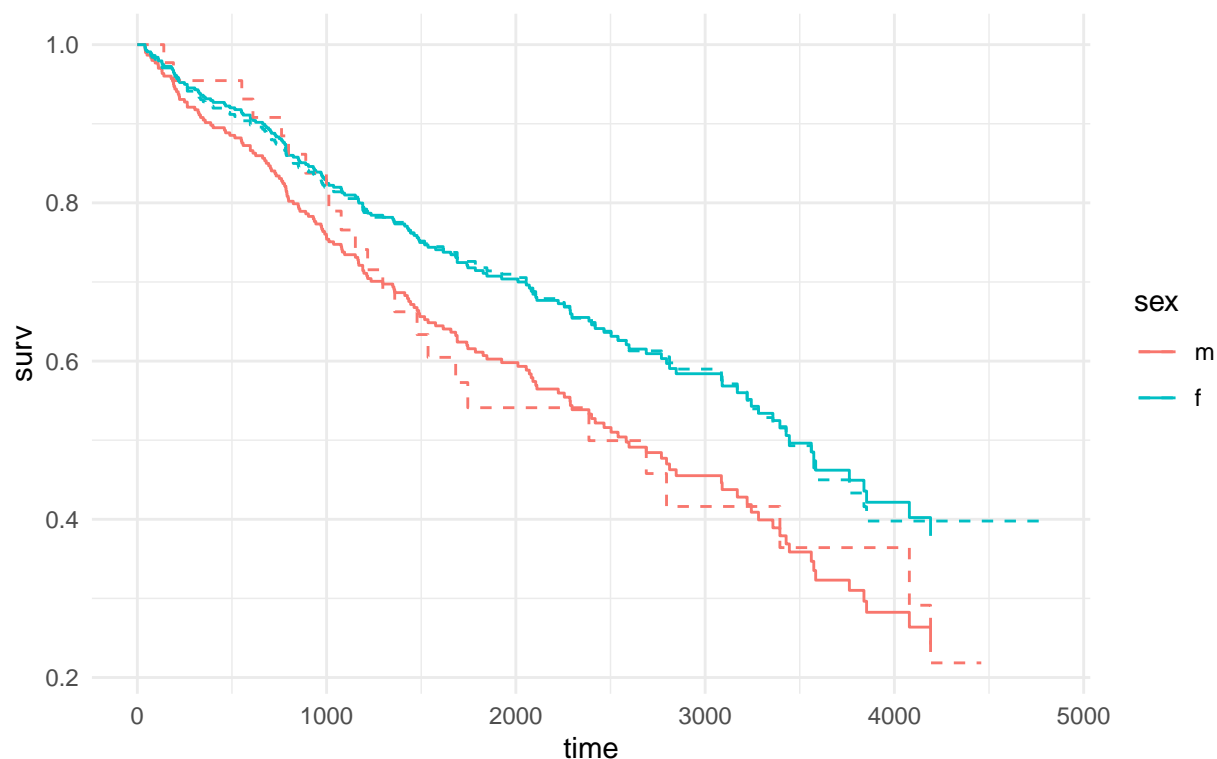
The dashed line is the Observed survival function

From the plot above, we can see that the fitted survival function of female group is pretty close to the observed survival function. However, the fitted survival function of the male group cross the observed survival function. That means the proportional assumption does not hold.

In order to investigate the interaction of the continuous variables with time, I included additional variables `albumin`, `bili`, `ast`, `copper` and `protime` and there interaction with time (in log scale) into the proportional hazard model. I choose those parameter based on the stepwise results from the last homework.

Table 2: Regresion Coefficients Estimates of the Cox Model with Time Interactions

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
sexf	-0.5014	6.057000e-01	0.3276	-1.5306	0.1259
albumin	22.9591	9.354306e+09	2.8063	8.1813	0.0000
bili	-1.2217	2.947000e-01	0.2493	-4.9013	0.0000
ast	0.0241	1.024400e+00	0.0208	1.1580	0.2469
copper	0.0315	1.032000e+00	0.0127	2.4813	0.0131
protime	10.3510	3.128905e+04	1.1481	9.0158	0.0000
albumin:log(time)	-3.2599	3.840000e-02	0.3893	-8.3737	0.0000
bili:log(time)	0.1801	1.197300e+00	0.0362	4.9787	0.0000
ast:log(time)	-0.0032	9.968000e-01	0.0028	-1.1576	0.2470
copper:log(time)	-0.0040	9.960000e-01	0.0018	-2.2024	0.0276
protime:log(time)	-1.3975	2.472000e-01	0.1599	-8.7426	0.0000

From the regression summary above, we can see that except `ast`, all the selected variable has a significant effect and interaction with time.
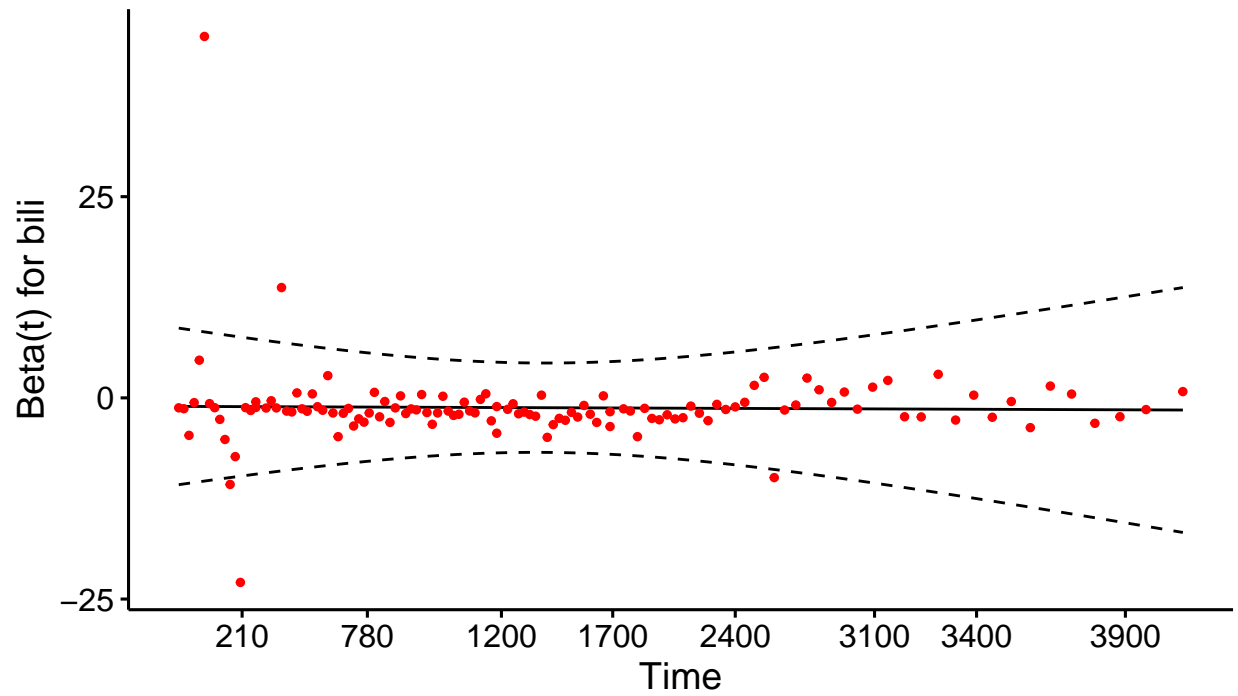
4

I plot the Schoenfeld residual of the fitted model with two continuous covariates, `albumin` and `copper`, below.

Global Schoenfeld Test p: 1.102e−39

Schoenfeld Individual Test p: 0
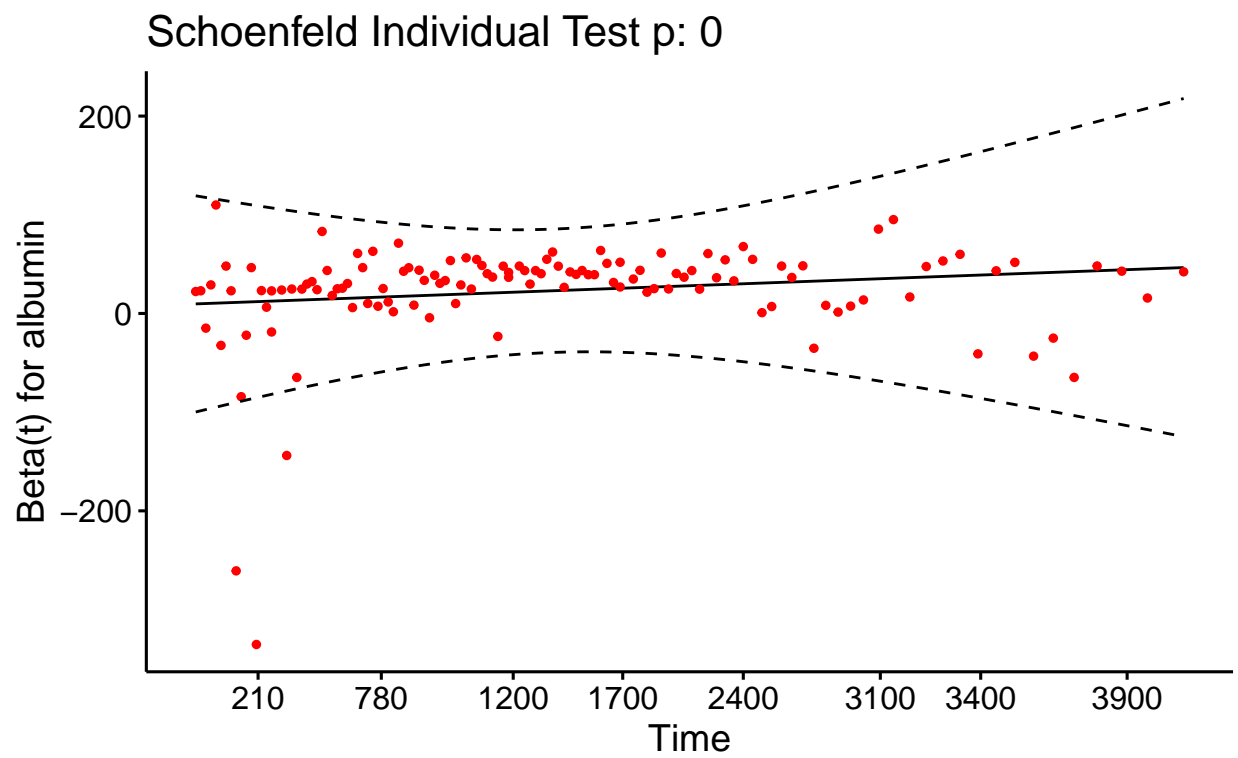
Global Schoenfeld Test p: 1.102e−39

Schoenfeld Individual Test p: 0.0465

From the above plot and the Schoenfeld individual test p-value, we can see that both the residual plots has a non-zero slope regression line, and the p-value of both tests are less than 0.05, which means both the covariates do not meet the PH assumption.

Appendix: Code for this report

```r
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
library(tidyverse)
library(knitr)
library(kableExtra)
library(survival)
heart_dat = readxl::read_excel("Datasets.xlsx", sheet = "Heartdata", col_names = FALSE)
colnames(heart_dat) = heart_dat[104,]
heart_dat = heart_dat[1:103,]
heart_dat$WAITING_TIME_FOR_TRANSPLANT = ifelse(is.na(heart_dat$WAITING_TIME_FOR_TRANSPLANT), 0, heart_da
heart_dat$AGE = as.numeric(heart_dat$AGE)
heart_dat$WAITING_TIME_FOR_TRANSPLANT = as.numeric(heart_dat$WAITING_TIME_FOR_TRANSPLANT)
heart_dat$AGE_T = heart_dat$AGE + ((heart_dat$WAITING_TIME_FOR_TRANSPLANT)/365.25)

heart_dat$TRANSPLANT_STATUS = as.factor(heart_dat$TRANSPLANT_STATUS)
heart_dat$SURVIVAL_STATUS = as.numeric(heart_dat$SURVIVAL_STATUS)
heart_dat$SURVIVAL_TIME = as.numeric(heart_dat$SURVIVAL_TIME)
heart_dat$PATIENT_ID = as.integer(heart_dat$PATIENT_ID)
heart.fit = coxph(Surv(SURVIVAL_TIME, SURVIVAL_STATUS==1) ~ TRANSPLANT_STATUS + AGE_T, data = heart_dat
heart.sum = summary(heart.fit)
rownames(heart.sum$coefficients) = c("Transplant Status = 1", "Age(t)")
heart.sum$coefficients %>% kable("latex", digits = 4, escape = F, booktabs = T, caption = "Regresion Co
library(ggsurvfit)
library(survminer)
pbc_survfit = survfit(Surv(time, status==2)~ sex, data = pbc)
pbc_survfit_log = survfit(Surv(log(time+1), status==2)~ sex, data = pbc)
#plot(pbc_survfit, col=c("black", "red"))
#plot(pbc_survfit, col=c("black", "red"), fun="cloglog")
splots <- list()
splots[[1]] <- ggsurvplot(pbc_survfit, data = pbc, risk.table = FALSE, ggtheme = theme_minimal(), conf.
splots[[2]] <- ggsurvplot(pbc_survfit_log, data = pbc, fun = "cloglog", risk.table = FALSE, xlab = "log

splots[[1]]$plot + labs(title = "Product Limit Survival Estimates")
splots[[2]]$plot + labs(title = "Log of Negative Log of Estimated Survival Function")
#arrange_ggsurvplots(splots, print = TRUE,
#  ncol = 2, nrow = 1, risk.table.height = 0.2)
#pbc_survfit %>%
#  ggsurvfit() +
#  labs(x = "Time",
#       y = "Overall survival probability") +
#  add_confidence_interval() +
#  add_risktable()
obsfit_plots = list()
obsfit_plots[[1]] = ggsurvplot(pbc_survfit, data = pbc, risk.table = FALSE, ggtheme = theme_minimal())
obsfit_plots[[2]] = ggadjustedcurves(coxph(Surv(time, status==2)~ sex, data = pbc), variable = "sex", da

cox_fit_surv_dat = obsfit_plots[[2]]$data
cox_fit_surv_dat$sex = cox_fit_surv_dat$variable

km_fit_surv_dat  = obsfit_plots[[1]]$plot$data
```

```
p <- ggplot(cox_fit_surv_dat, aes(x = time, y = surv, group = sex, color = sex)) + geom_step()
p + geom_step(data = km_fit_surv_dat, aes(x = time, y = surv, group = sex, color = sex), lty = 2) + labs
pbc_interaction_fit = coxph(Surv(time, status==2)~ sex + albumin + bili + ast + copper + protime + log(
summary(pbc_interaction_fit)$coefficients%>% kable("latex", digits = 4, escape = F, booktabs = T, capti
ggcoxzph(cox.zph(pbc_interaction_fit), var = c("albumin"),df = 2, nsmo = 1000)
ggcoxzph(cox.zph(pbc_interaction_fit), var = c("bili"),df = 2, nsmo = 1000)
```