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, 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.

```
## Call:
## coxph(formula = Surv(start, stop, event) ~ transplant * age_trans,
##     data = heart_wait_dat, id = id)
##
##   n= 172, number of events= 75
##
##                          coef exp(coef) se(coef)      z Pr(>|z|)
## transplant           -0.93994   0.39065  1.31868 -0.713    0.476
## age_trans            -0.01247   0.98760  0.04218 -0.296    0.767
## transplant:age_trans  0.02105   1.02127  0.02766  0.761    0.447
##
##                      exp(coef) exp(-coef) lower .95 upper .95
## transplant              0.3907     2.5598   0.02947     5.179
## age_trans               0.9876     1.0126   0.90924     1.073
## transplant:age_trans    1.0213     0.9792   0.96739     1.078
##
## Concordance= 0.538  (se = 0.038 )
## Likelihood ratio test= 2.64  on 3 df,   p=0.5
## Wald test            = 2.52  on 3 df,   p=0.5
## Score (logrank) test = 2.51  on 3 df,   p=0.5
```

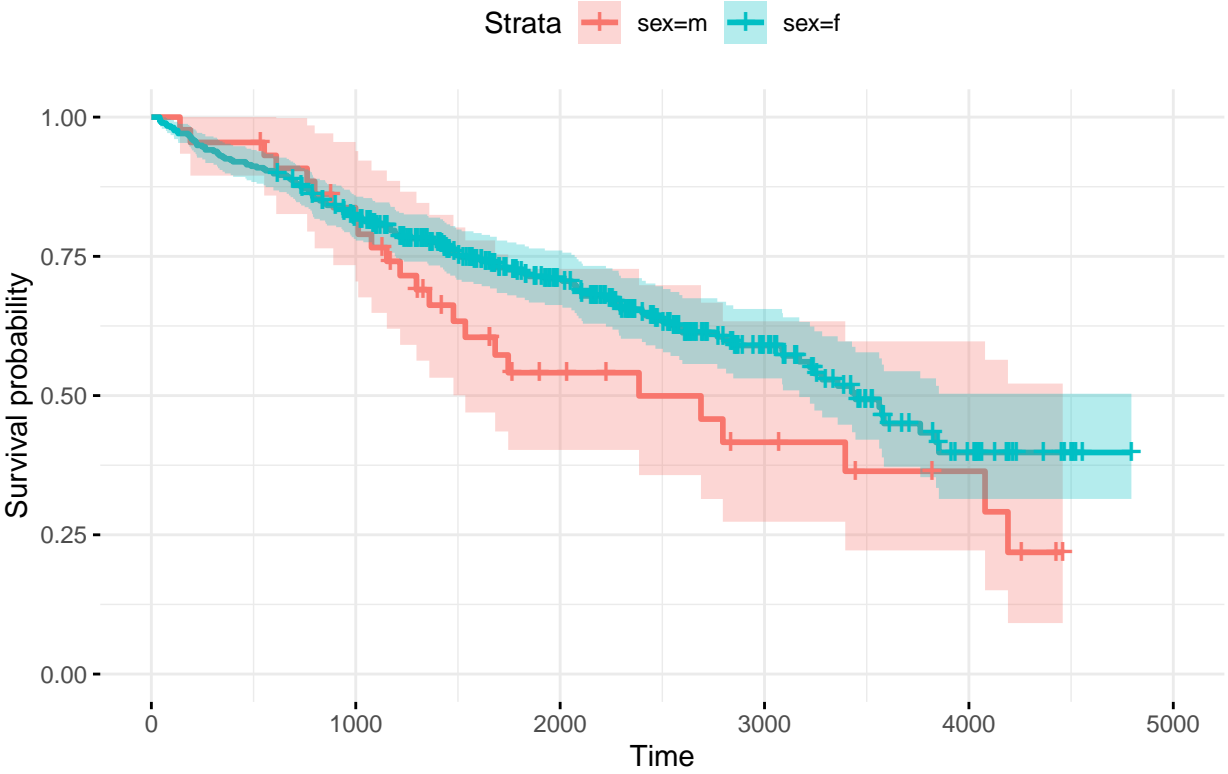*[handwritten annotations in margin: "baseline age + transplant · age-trans", "trans + age (...", "age(t) = ∫ age+trans / age+w", "trans =1"]*
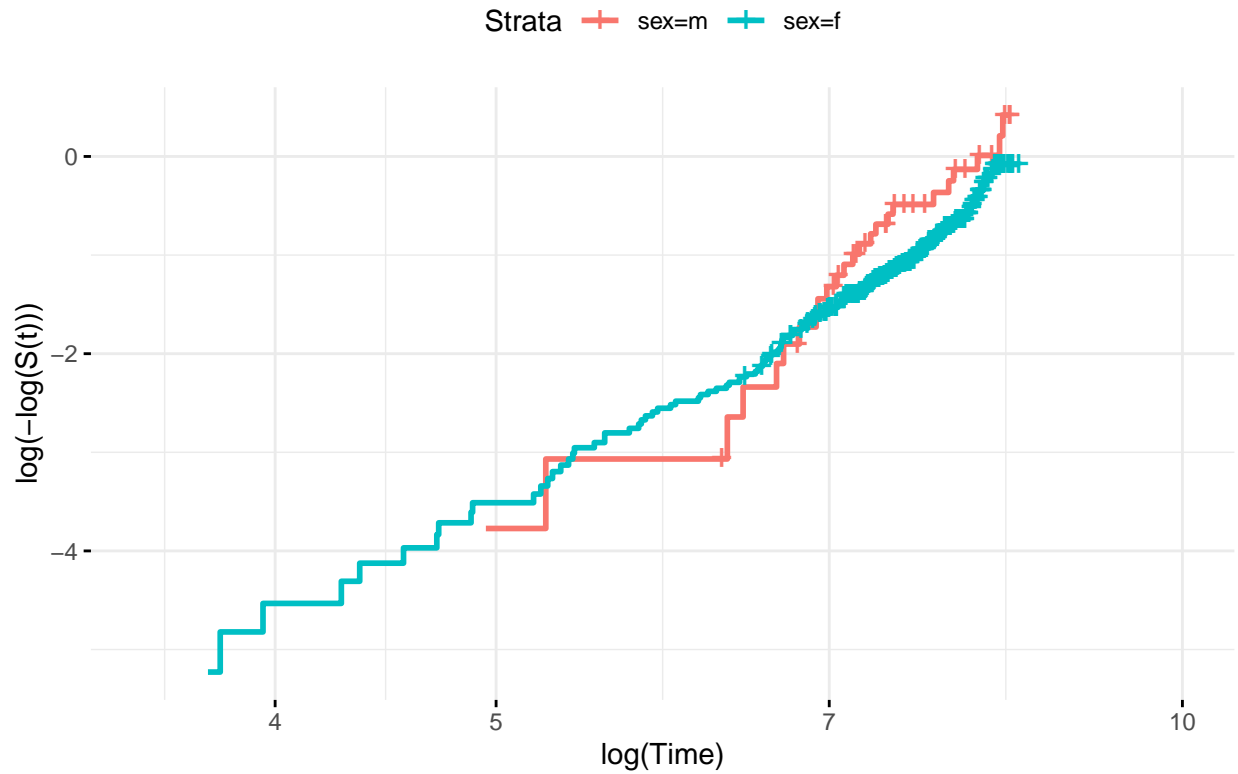
Problem 2

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
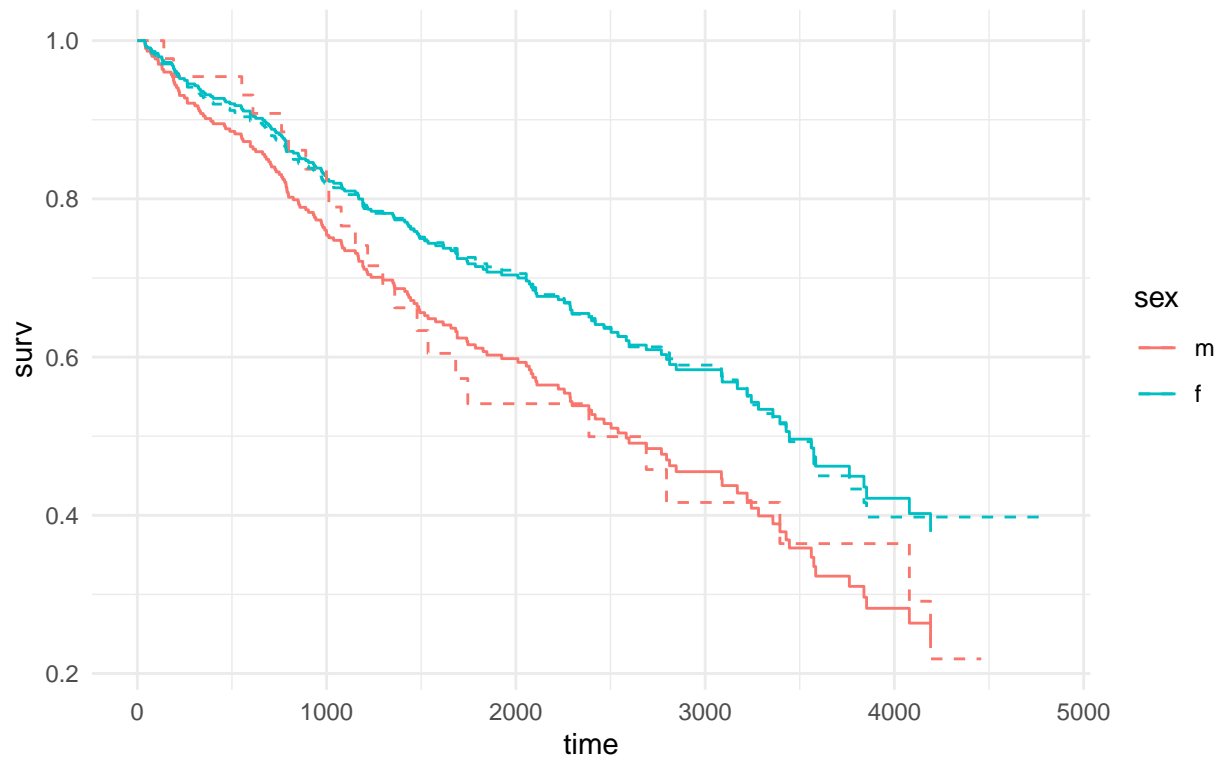
The dashed line is the Observed survival function

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 `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.

```
## Call:
## coxph(formula = Surv(time, status == 2) ~ albumin + bili + ast +
##     copper + protime + log(time):albumin + log(time):bili + log(time):ast +
##     log(time):copper + log(time):protime, data = pbc)
##
##   n= 310, number of events= 124
##    (108 observations deleted due to missingness)
##
##                         coef  exp(coef)   se(coef)       z Pr(>|z|)
## albumin            2.256e+01  6.294e+09  2.804e+00   8.047 8.47e-16 ***
## bili              -1.280e+00  2.782e-01  2.479e-01  -5.162 2.45e-07 ***
## ast                1.946e-02  1.020e+00  2.058e-02   0.946   0.3444
## copper             3.075e-02  1.031e+00  1.274e-02   2.414   0.0158 *
## protime            1.051e+01  3.676e+04  1.141e+00   9.210  < 2e-16 ***
## albumin:log(time) -3.204e+00  4.062e-02  3.890e-01  -8.236  < 2e-16 ***
## bili:log(time)     1.872e-01  1.206e+00  3.606e-02   5.192 2.08e-07 ***
## ast:log(time)     -2.642e-03  9.974e-01  2.749e-03  -0.961   0.3366
## copper:log(time)  -3.762e-03  9.962e-01  1.803e-03  -2.086   0.0370 *
## protime:log(time) -1.421e+00  2.416e-01  1.589e-01  -8.938  < 2e-16 ***
```

4

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## albumin             6.294e+09  1.589e-10 2.584e+07 1.533e+12
## bili                2.782e-01  3.595e+00 1.711e-01 4.522e-01
## ast                 1.020e+00  9.807e-01 9.793e-01 1.062e+00
## copper              1.031e+00  9.697e-01 1.006e+00 1.057e+00
## protime             3.676e+04  2.721e-05 3.924e+03 3.443e+05
## albumin:log(time) 4.062e-02  2.462e+01 1.895e-02 8.706e-02
## bili:log(time)    1.206e+00  8.293e-01 1.124e+00 1.294e+00
## ast:log(time)     9.974e-01  1.003e+00 9.920e-01 1.003e+00
## copper:log(time)  9.962e-01  1.004e+00 9.927e-01 9.998e-01
## protime:log(time) 2.416e-01  4.139e+00 1.769e-01 3.299e-01
##
## Concordance= 0.992  (se = 0.002 )
## Likelihood ratio test= 804.9  on 10 df,   p=<2e-16
## Wald test            = 143.4  on 10 df,   p=<2e-16
## Score (logrank) test = 895.2  on 10 df,   p=<2e-16
```

From the regression summary above, we can see that except `ast`, all the selected variable has a significant effcet and interaction with time.
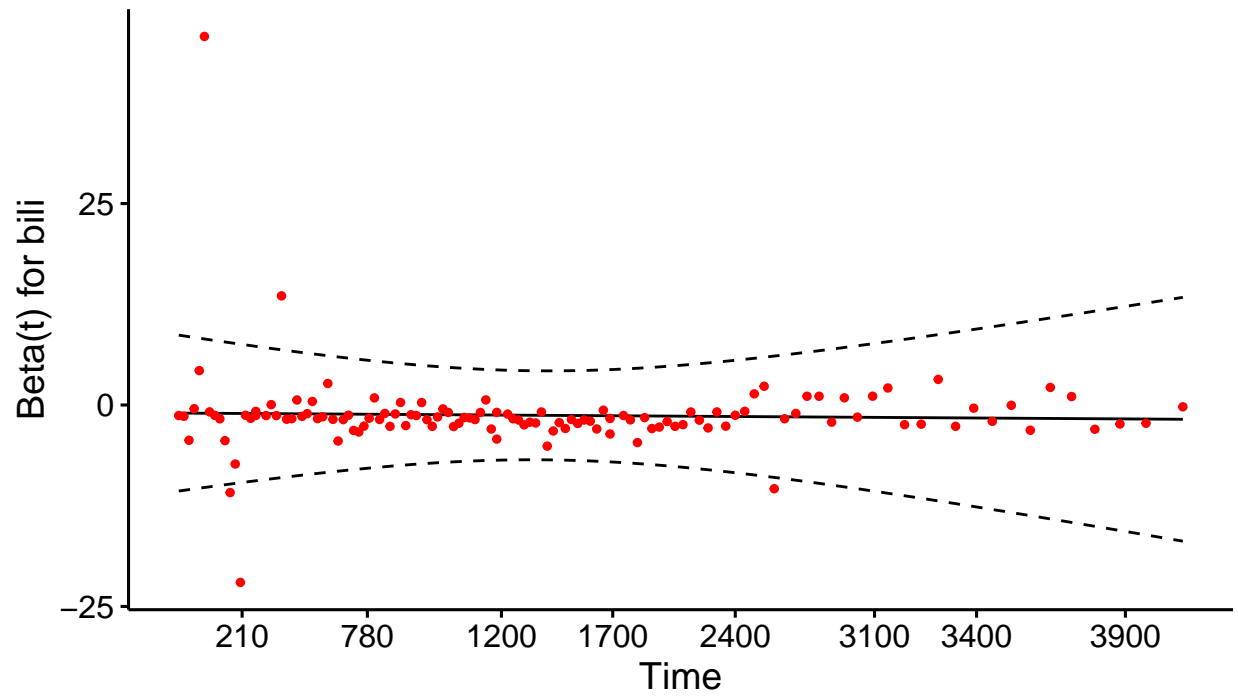
I plot the Schoenfeld residual of the fitted model with two continuous covariates, `albumin` and `copper`, below.

Global Schoenfeld Test p: 2.21e−39

Schoenfeld Individual Test p: 0
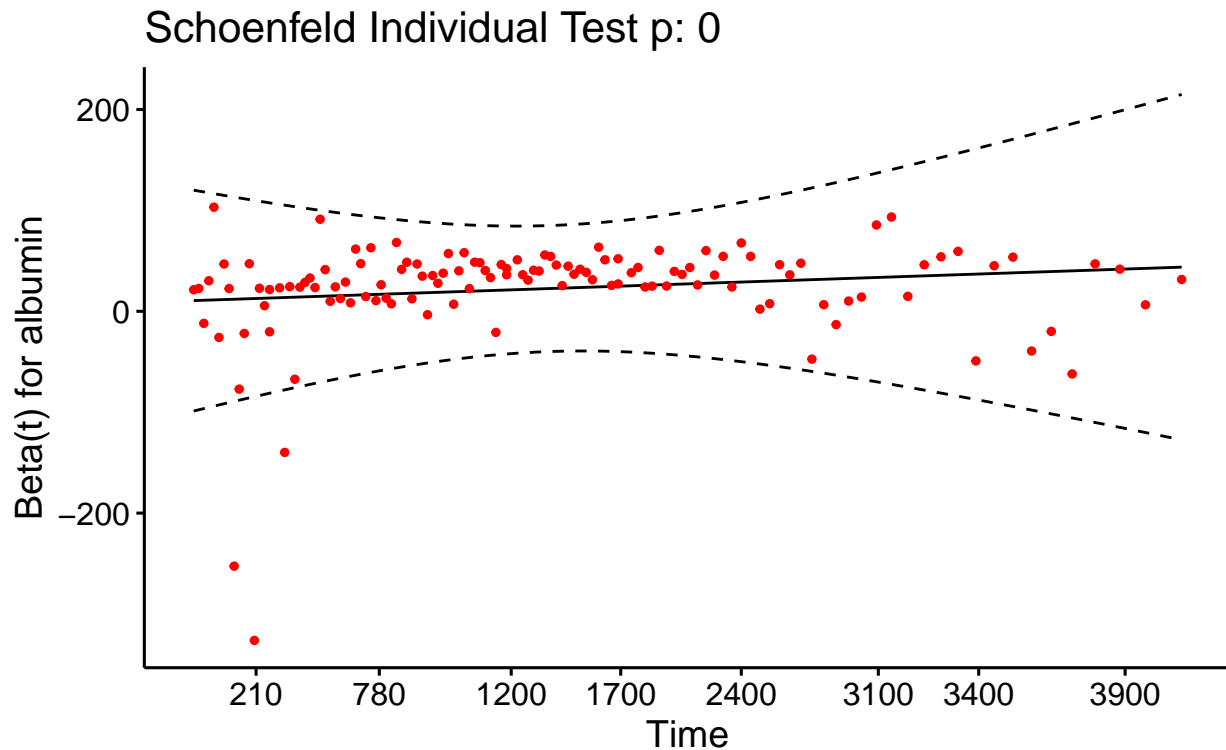
Global Schoenfeld Test p: 2.21e−39

Schoenfeld Individual Test p: 0.0419

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.

Reference:

1. Munshi, N. C., Anderson, L. D., Jr, Shah, N., Madduri, D., Berdeja, J., Lonial, S., Raje, N., Lin, Y., Siegel, D., Oriol, A., Moreau, P., Yakoub-Agha, I., Delforge, M., Cavo, M., Einsele, H., Goldschmidt, H., Weisel, K., Rambaldi, A., Reece, D., Petrocca, F., . . . San-Miguel, J. (2021). Idecabtagene Vicleucel in Relapsed and Refractory Multiple Myeloma. The New England journal of medicine, 384(8), 705–716. https://doi.org/10.1056/NEJMoa2024850

Appendix: Code for this report

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
library(tidyverse)
library(knitr)
library(kableExtra)
library(survival)
heart_dat = heart
heart_wait = heart_dat %>% select(start, stop, transplant, id)  %>% mutate(interval = stop - start) %>%
heart_wait_dat = merge(heart_dat, heart_wait, by=c("start","stop","id","transplant"))
heart_wait_dat$age = heart_wait_dat$age + 48
heart_wait_dat$age_trans = heart_wait_dat$age + (heart_wait_dat$wait_time/365.25)
heart_wait_dat$transplant = as.numeric(heart_wait_dat$transplant)
heart.fit = coxph(Surv(start, stop, event) ~ transplant*age_trans, data = heart_wait_dat, id = id)
summary(heart.fit)
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)~ albumin + bili + ast + copper + protime + log(time):a
summary(pbc_interaction_fit)
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)
```