PH model checking

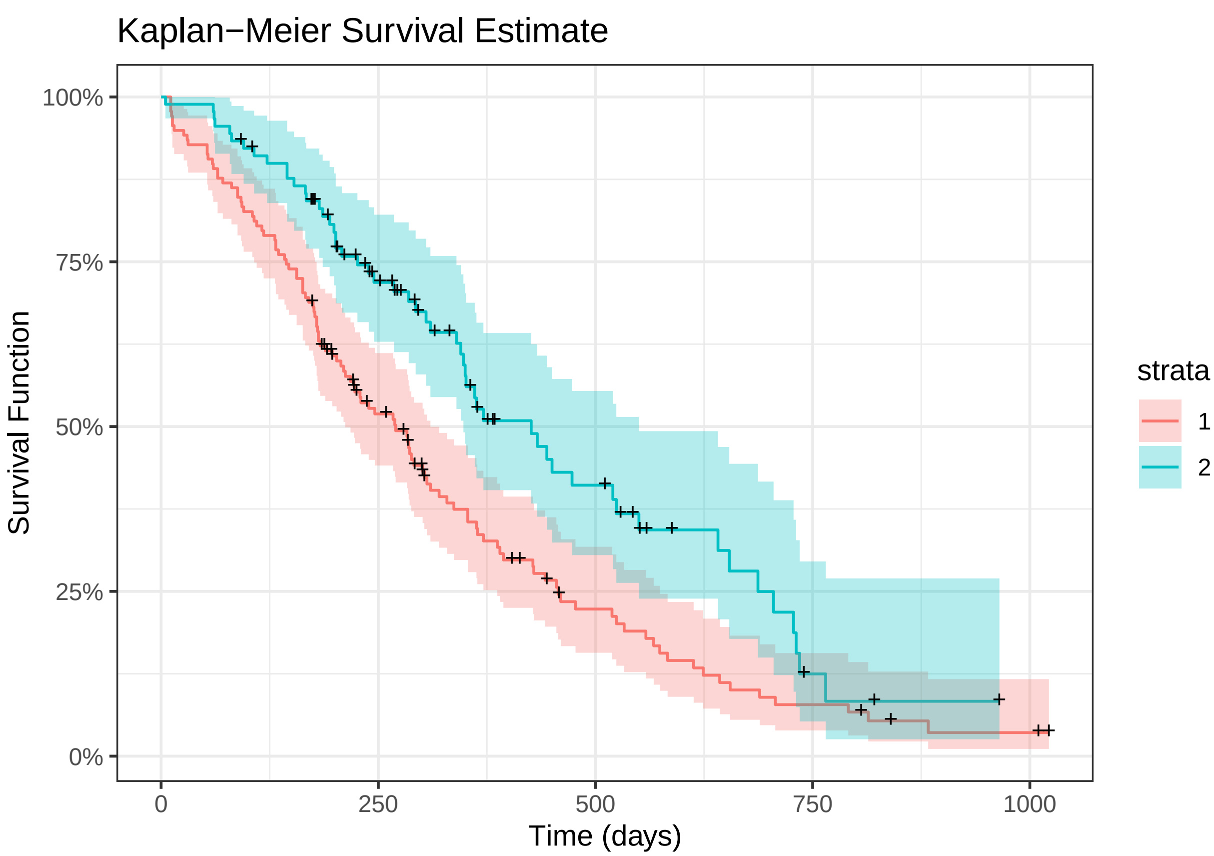
2022-11-27

Import data

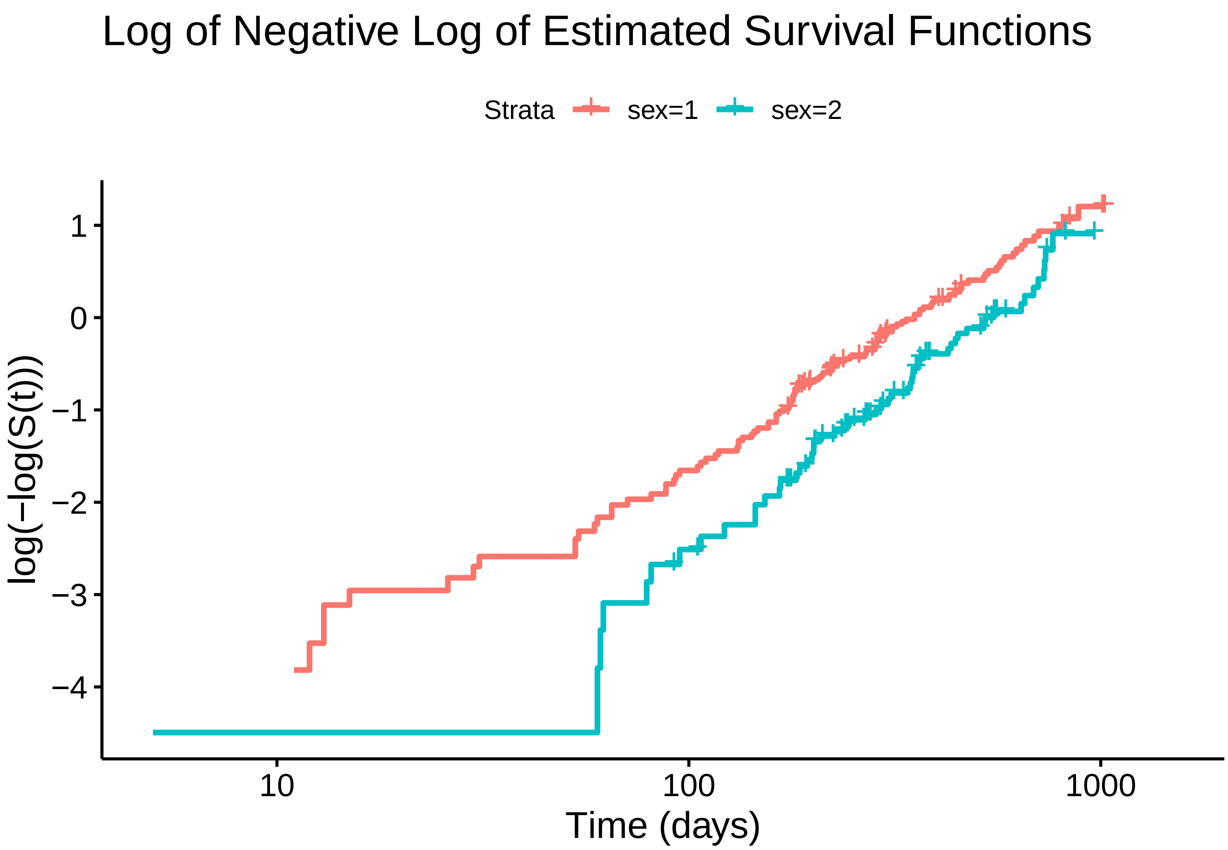
```
dat_lung_raw <- survival::lung
dat_lung <- dat_lung_raw %>%
  mutate_at(c(1, 3, 5, 6), .funs = ~as.factor(.))
```

Model checking I

```
# --- sex ---
# km plot
fit_km_sex <- survfit(Surv(time, status == 2) ~ sex, dat_lung)
autoplot(fit_km_sex) + theme_bw() +
  labs(x = "Time (days)", y = "Survival Function",
       title = "Kaplan-Meier Survival Estimate")
```

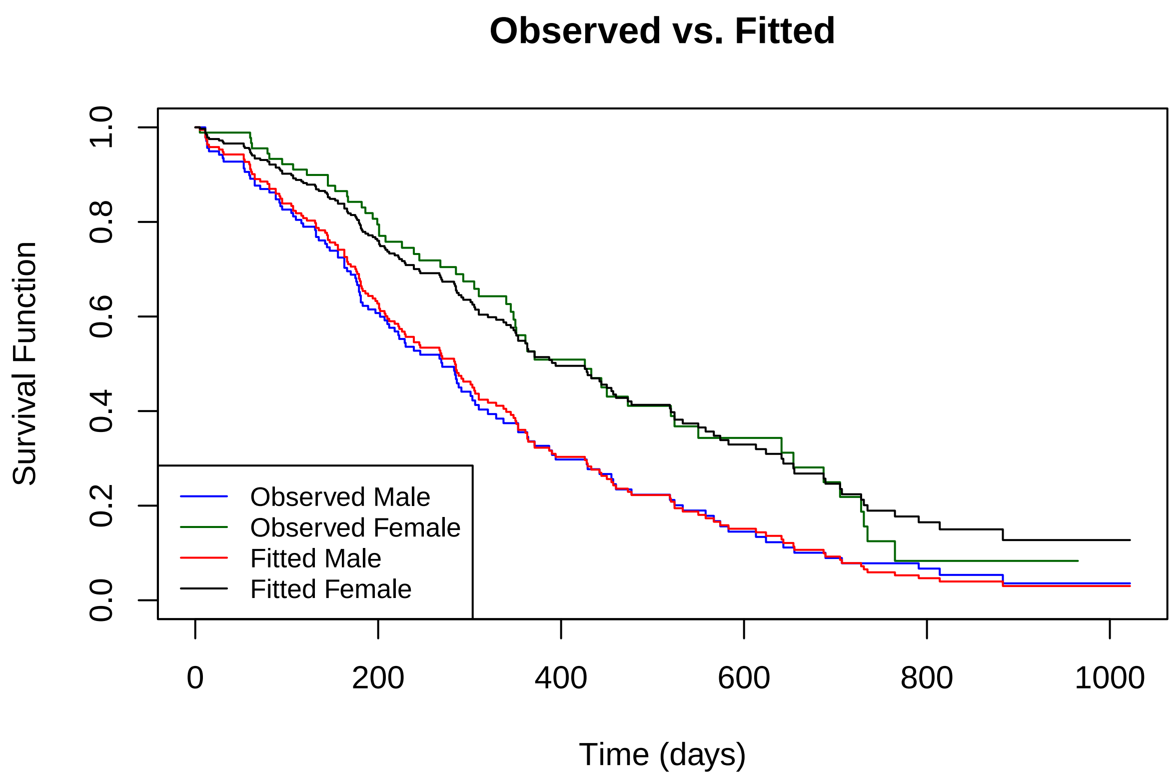


```r
# loglog vs. time
ggsurvplot(fit_km_sex, fun = "cloglog", xlim = c(5, 1500),
           xlab = "Time (days)",
           title = "Log of Negative Log of Estimated Survival Functions")
```




```r
# observed vs. fitted
fit_ph_sex <- coxph(Surv(time, status == 2) ~ sex, dat_lung)

plot(fit_km_sex, col = c("blue", "darkgreen"),
     xlab = "Time (days)", ylab = "Survival Function",
     main = "Observed vs. Fitted")
lines(survfit(fit_ph_sex, newdata = data.frame(sex = as.factor(1))), # male
      col = "red", conf.int = FALSE)
lines(survfit(fit_ph_sex, newdata = data.frame(sex = as.factor(2))), # female
      col = "black", conf.int = FALSE)
legend("bottomleft", legend = c("Observed Male", "Observed Female",
                                "Fitted Male", "Fitted Female"),
       col = c("blue", "darkgreen", "red", "black"), lty = 1, cex = .8)
```



Interpretation: The above two figures demonstrate the proportional hazards assumption is hold given there is only one indicator variable `sex` in the model.

Model checking II

```r
# --- to be updated ---
# stepwise selection
stepw_res <- StepReg::stepwiseCox(Surv(time, status == 2) ~., dat_lung %>% na.omit(),
                                  selection = "bidirection", select = "AIC")
stepw_res
```

```
## $`Basic Information`
##
## 1        Response Variable =  Surv(time, status == 2)
## 2        Included Variable =                     NULL
## 3         Selection Method =              bidirection
## 4         Select Criterion =                      AIC
## 5 Multicollinearity Terms =                    efron
## 6                   Method =                     NULL
##
## $`Variable Class`
##        class                           variable
## 1    factor:                   inst sex ph.ecog
## 2 nmatrix.2:            Surv(time, status == 2)
## 3   numeric: age ph.karno pat.karno meal.cal wt.loss
##
## $Process
```

```
##   Step EnteredEffect RemovedEffect DF NumberIn              AIC
## 2    1       ph.ecog                3        3 1008.58427913433
## 3    2           sex                1        4 1003.84247093824
## 4    3       wt.loss                1        5 1003.06920656013
## 5    4      ph.karno                1        6 1002.15358992223
##
## $Variables
## [1] "ph.ecog"  "sex"      "wt.loss"  "ph.karno"
##
## $Coefficients
##                 coef  exp(coef)    se(coef)         z     Pr(>|z|)
## ph.ecog2  0.65012440  1.9157791 0.280695973  2.316116 0.0205519425
## ph.ecog3  1.67694368  5.3491822 0.441824118  3.795501 0.0001473459
## ph.ecog4  2.88359035 17.8783476 1.121914073  2.570242 0.0101627579
## sex2     -0.56468797  0.5685375 0.199666717 -2.828153 0.0046817462
## wt.loss  -0.01279297  0.9872885 0.007676834 -1.666438 0.0956262406
## ph.karno  0.01853809  1.0187110 0.011153914  1.662025 0.0965077398
```

```
# interaction
fit_ph_1 <- coxph(Surv(time, status == 2) ~ sex + wt.loss + ph.karno +
                    sex * log(time) + wt.loss * log(time) + ph.karno * log(time),
                  dat_lung %>% na.omit())
summary(fit_ph_1)
```

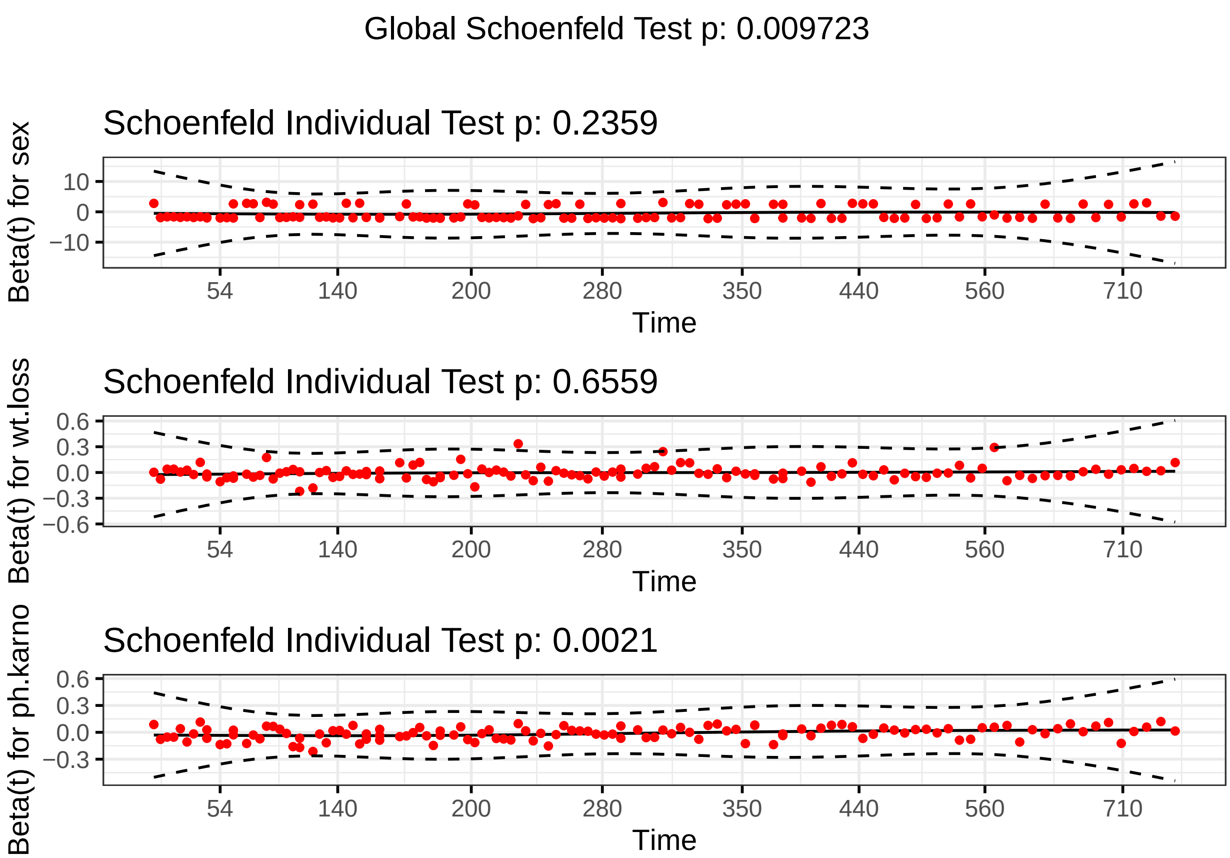
```
## Call:
## coxph(formula = Surv(time, status == 2) ~ sex + wt.loss + ph.karno +
##     sex * log(time) + wt.loss * log(time) + ph.karno * log(time),
##     data = dat_lung %>% na.omit())
##
##   n= 167, number of events= 120
##
##                           coef  exp(coef)   se(coef)      z Pr(>|z|)
## sex2                -1.675e+00  1.874e-01  3.603e-01 -4.648 3.35e-06 ***
## wt.loss              8.469e-02  1.088e+00  1.208e-02  7.009 2.40e-12 ***
## ph.karno             7.232e-03  1.007e+00  1.390e-02  0.520   0.6028
## log(time)           -1.786e+02  2.651e-78  2.613e+01 -6.835 8.18e-12 ***
## sex2:log(time)       2.974e-01  1.346e+00  6.319e-02  4.706 2.52e-06 ***
## wt.loss:log(time)   -1.540e-02  9.847e-01  2.139e-03 -7.201 5.96e-13 ***
## ph.karno:log(time) -4.522e-03  9.955e-01  2.440e-03 -1.853   0.0638 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                    exp(coef) exp(-coef)  lower .95 upper .95
## sex2               1.874e-01  5.337e+00  9.249e-02 3.797e-01
## wt.loss            1.088e+00  9.188e-01  1.063e+00 1.114e+00
## ph.karno           1.007e+00  9.928e-01  9.802e-01 1.035e+00
## log(time)          2.651e-78  3.773e+77 1.511e-100 4.651e-56
## sex2:log(time)     1.346e+00  7.427e-01  1.190e+00 1.524e+00
## wt.loss:log(time)  9.847e-01  1.016e+00  9.806e-01 9.889e-01
## ph.karno:log(time) 9.955e-01  1.005e+00  9.907e-01 1.000e+00
##
## Concordance= 1  (se = 0 )
## Likelihood ratio test= 950.1  on 7 df,   p=<2e-16
## Wald test            = 195.2  on 7 df,   p=<2e-16
```

```
## Score (logrank) test = 398.7  on 7 df,   p=<2e-16
```

```
# residual
fit_ph_2 <- coxph(Surv(time, status == 2) ~ sex + wt.loss + ph.karno, dat_lung %>% na.omit())
test_ph <- cox.zph(fit_ph_2)
test_ph
```

```
##           chisq df      p
## sex       1.405  1 0.2359
## wt.loss   0.199  1 0.6559
## ph.karno  9.488  1 0.0021
## GLOBAL   11.406  3 0.0097
```

```
ggcoxzph(test_ph, ggtheme = theme_bw())
```



Description and Interpretation:

- Stepwise procedure was performed to select variables from the original dataset and finalize **4** variables including ph.ecog, sex, wt.loss, ph.karno;
- Test interaction for proportionality given **3** selected variables sex, wt.loss, ph.karno, the final result shows that only the interaction of `ph.karno` with `log(time)` is non-significant (i.e., greater than 0.05). So we conclude that the proportionality assumption for `ph.karno` is violated. The Schoenfeld residual of the fitted model agrees with such conclusion.

Time-varying model

Ways to fit a time-varying model…