

PH model checking

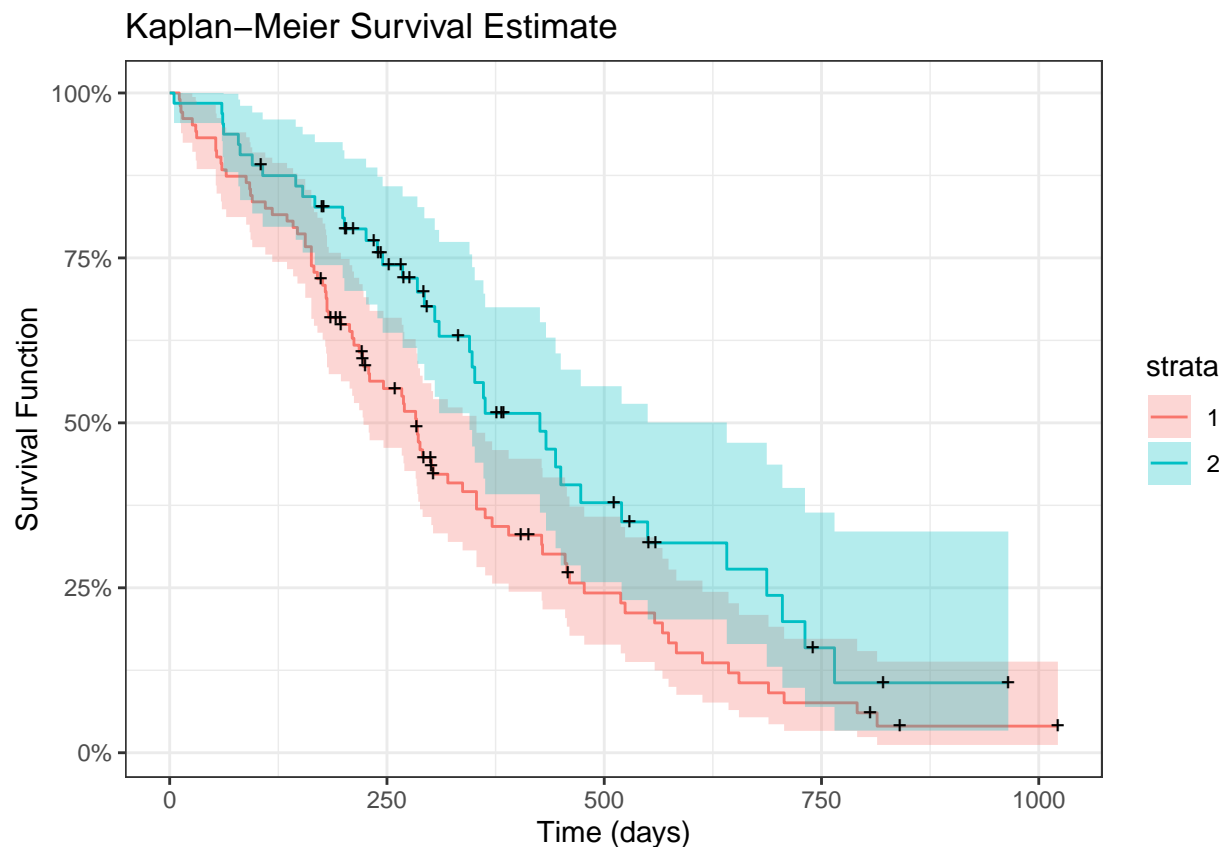
Qihang Wu

2022-December-02

Model checking I (graphical methods)

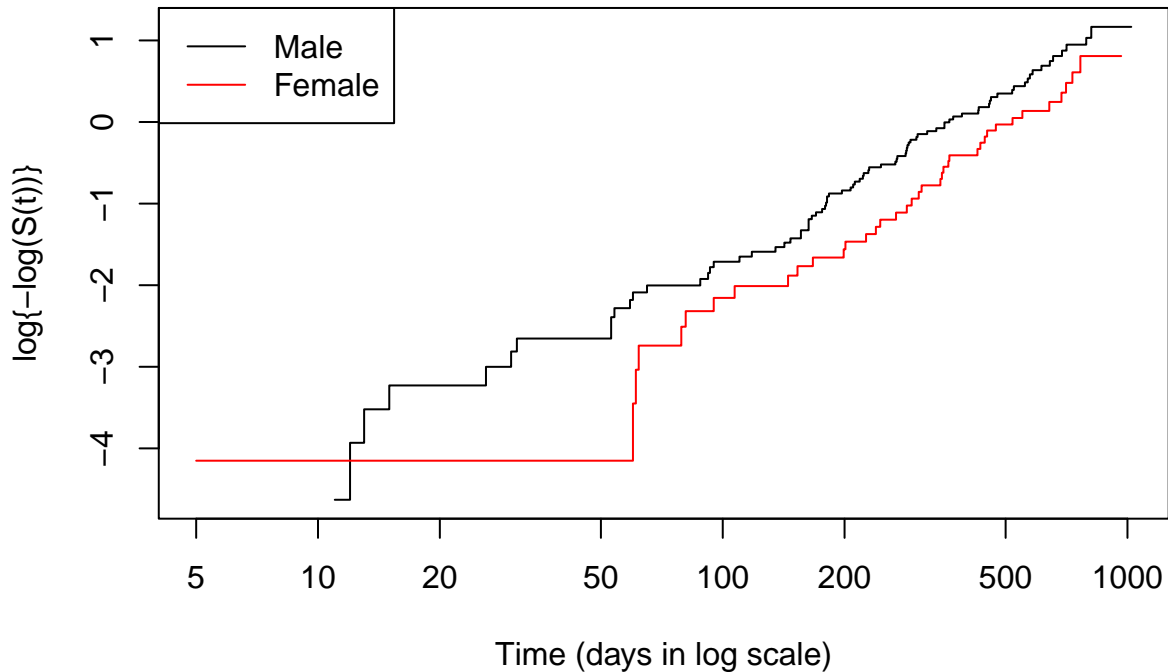
```
# import data
dat_lung_raw <- survival::lung
dat_lung <- dat_lung_raw %>%
  mutate_at(c(1, 3, 5, 6), .funs = ~as.factor(.)) %>% na.omit()

# --- 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")
```



```
# loglog vs. log time
# png("ph_checking_1.png", width = 500, height = 400)
plot(fit_km_sex, fun = "cloglog", col = c("black", "red"),
     xlab = "Time (days in log scale)", ylab = "log{-log(S(t))}",
     main = "Log of Negative Log of Estimated Survival Functions")
legend("topleft", legend = c("Male", "Female"), col = c("black", "red"),
      lty = 1, cex = 1)
```

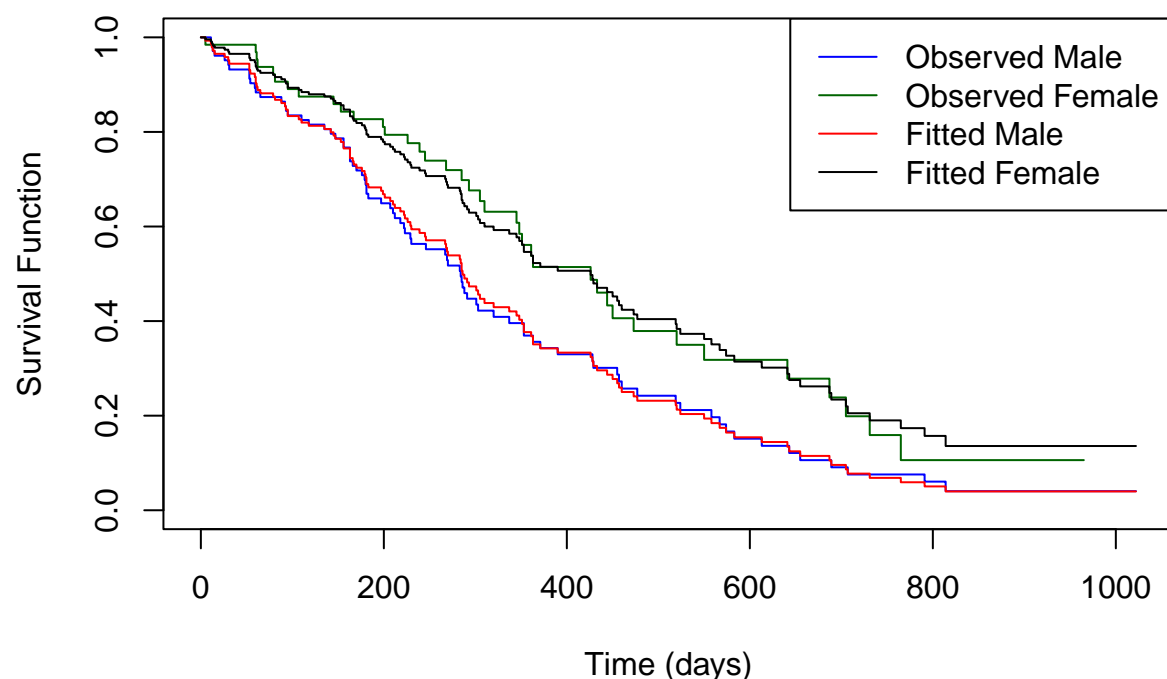
Log of Negative Log of Estimated Survival Functions



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

# png("ph_checking_2.png", width = 500, height = 400)
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("topright", legend = c("Observed Male", "Observed Female",
                              "Fitted Male", "Fitted Female"),
      col = c("blue", "darkgreen", "red", "black"), lty = 1, cex = 1)
```

Observed vs. Fitted



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

```
# --- to be updated ---
# stepwise selection
StepReg::stepwiseCox(Surv(time, status == 2) ~., dat_lung,
                     selection = "bidirection", select = "AIC")
```

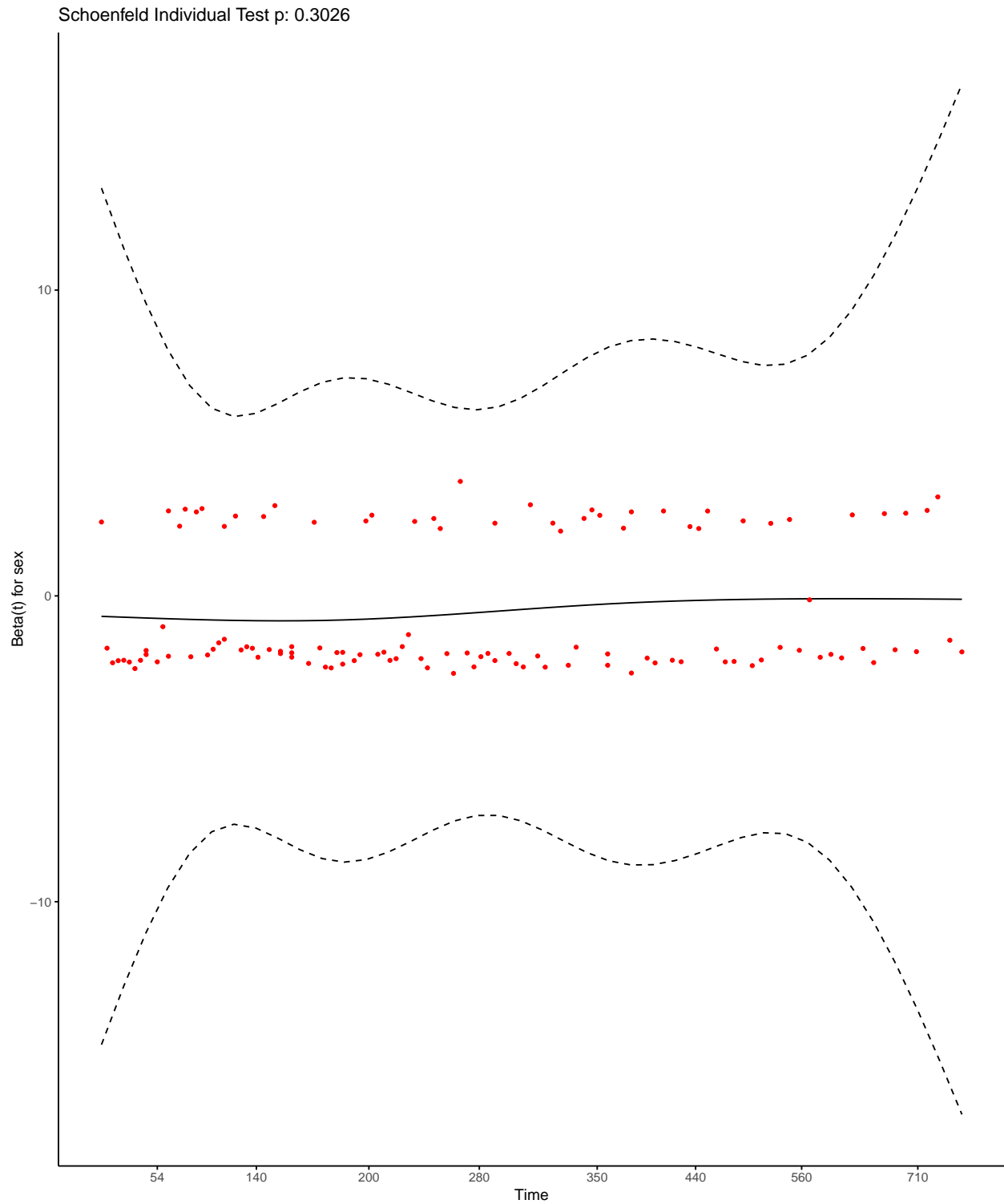
```
## $'Basic Information'
##
## 1      Response Variable =  Surv(time, status == 2)
## 2      Included Variable =                               NULL
## 3      Selection Method =                               bidirection
## 4      Select Criterion =                               AIC
## 5 Multicollinearity 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
```

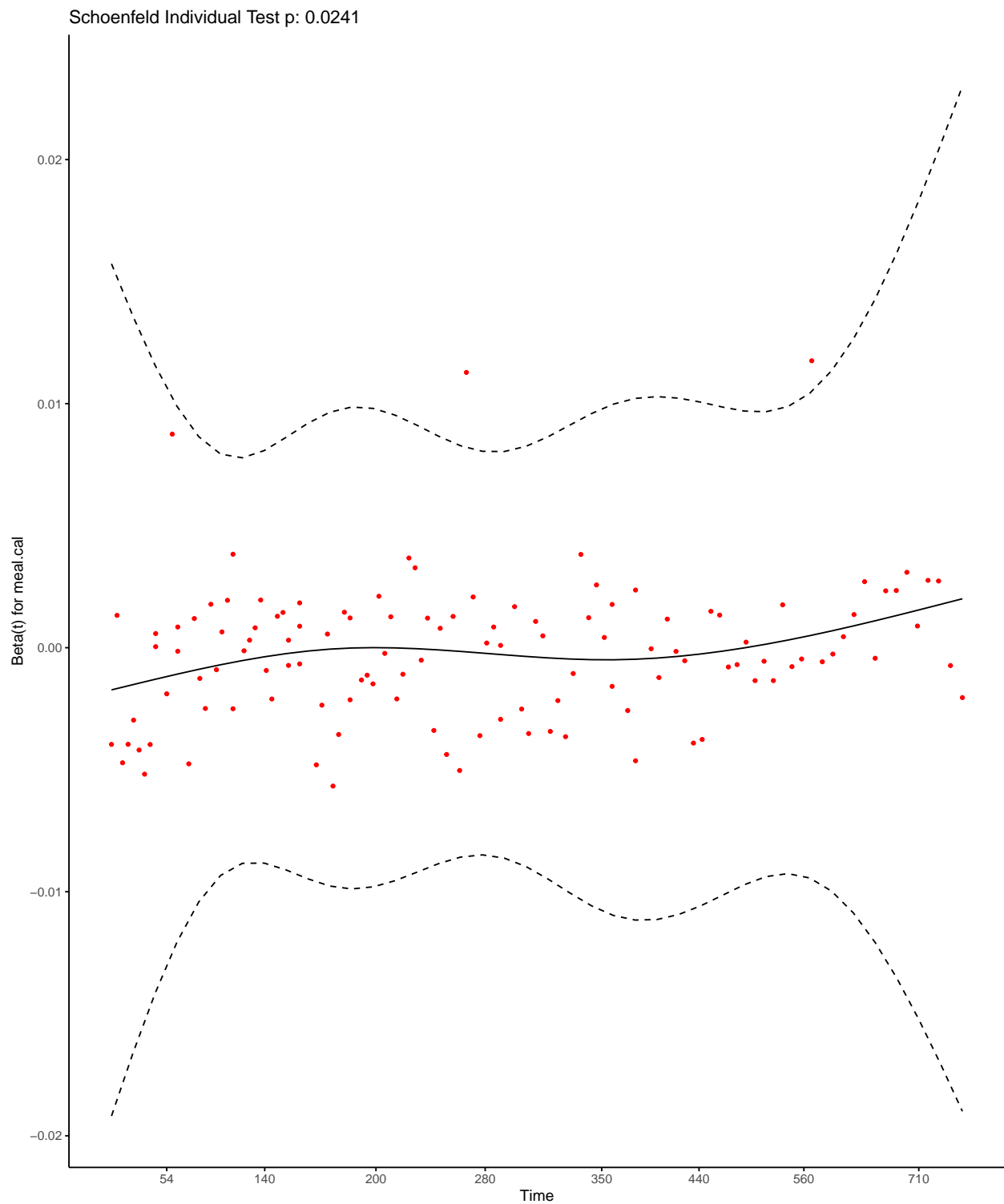
```
# interaction
fit_ph_1 <- coxph(Surv(time, status == 2) ~ sex + meal.cal + wt.loss +
                  sex * time + meal.cal * time + wt.loss * time,
                  dat_lung)
summary(fit_ph_1)
```

```
## Call:
## coxph(formula = Surv(time, status == 2) ~ sex + meal.cal + wt.loss +
##       sex * time + meal.cal * time + wt.loss * time, data = dat_lung)
##
##      n= 167, number of events= 120
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## sex2      -4.053e-01  6.668e-01  4.787e-01 -0.847  0.397
## meal.cal    2.345e-04  1.000e+00  4.447e-04  0.527  0.598
## wt.loss     1.041e-02  1.010e+00  1.619e-02  0.643  0.520
## time      -1.408e+00  2.446e-01  2.450e-01 -5.748 9.02e-09 ***
## sex2:time     6.079e-04  1.001e+00  1.583e-03  0.384  0.701
## meal.cal:time -7.809e-07  1.000e+00  1.778e-06 -0.439  0.661
## wt.loss:time  -2.643e-05  1.000e+00  6.428e-05 -0.411  0.681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## sex2          0.6668    1.4998    0.2609    1.7039
## meal.cal       1.0002    0.9998    0.9994    1.0011
## wt.loss        1.0105    0.9896    0.9789    1.0430
## time           0.2446    4.0885    0.1513    0.3953
## sex2:time       1.0006    0.9994    0.9975    1.0037
## meal.cal:time   1.0000    1.0000    1.0000    1.0000
## wt.loss:time    1.0000    1.0000    0.9998    1.0001
##
## Concordance= 1 (se = 0 )
## Likelihood ratio test= 973.8 on 7 df, p=<2e-16
## Wald test              = 34.96 on 7 df, p=1e-05
## Score (logrank) test = 184.4 on 7 df, p=<2e-16
```

```
# residual
fit_ph_2 <- coxph(Surv(time, status == 2) ~ sex + meal.cal + wt.loss, dat_lung)
test_ph <- cox.zph(fit_ph_2)
ggcoxzph(test_ph[1], ggtheme = theme_classic())
```



```
# ggsave("ph_checking_3.png", width = 6, height = 4)
ggcoxzph(test_ph[2], ggtheme = theme_classic())
```



```
# ggsave("ph_checking_4.png", width = 6, height = 4)
ggcoxzph(test_ph[3], ggtheme = theme_classic())
```

