Day9 exercise solutions

Ali Movasati, Isabelle Caroline Rose Cretton, Tristan Koning

Nov. 15th, 2024

```r
# Set global code chunk options
knitr::opts_chunk$set(
  echo = TRUE,
  warning = FALSE,
  message = FALSE,
  fig.width = 10,
  fig.height = 6
)
```

```r
# load required libraries
library("skimr")
library("dplyr")
library("magrittr")
library("ggplot2")
library("survival")
library("survminer")
library("gridExtra")

# define functions
`%notin%` <- Negate(`%in%`)
```

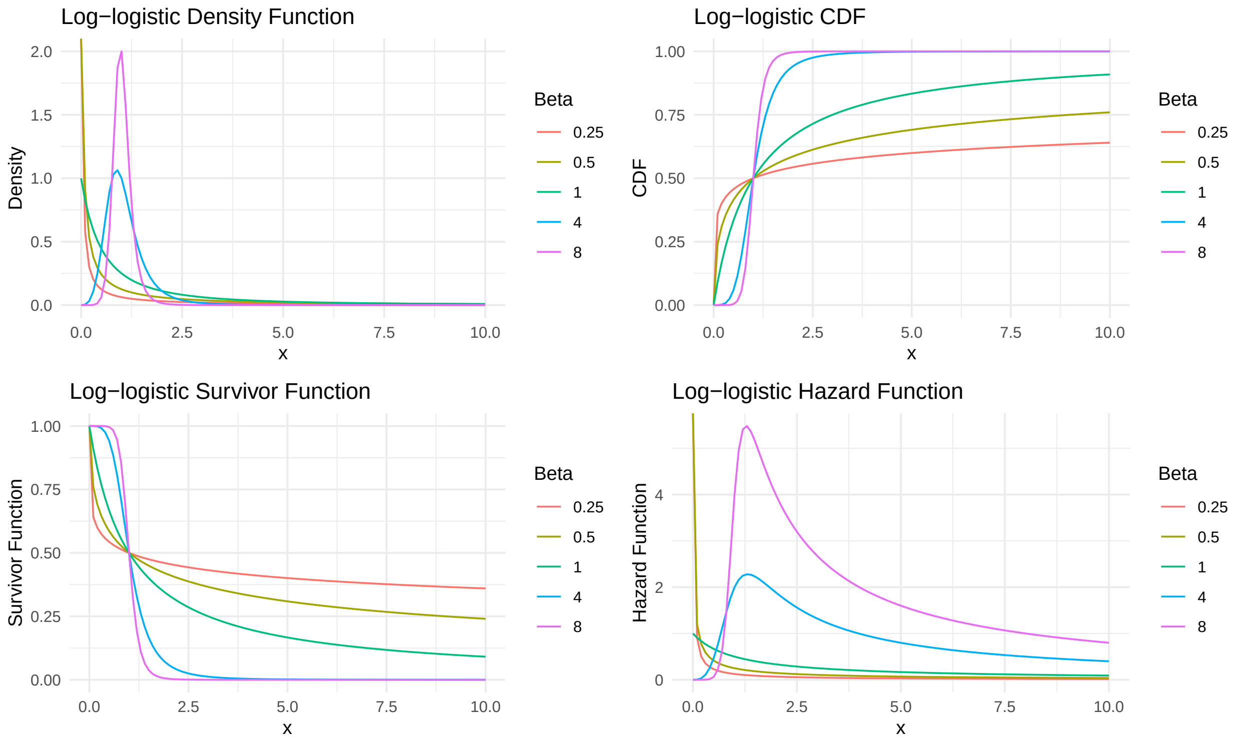
Problem 1

```r
log_logistic_density <- function(x, alpha, beta) {
    numerator <- (beta / alpha) * (x / alpha) ^ (beta - 1)
    denominator <- (1 + (x / alpha) ^ beta) ^ 2
    return(numerator / denominator)
}

log_logistic_cdf <- function(x, alpha, beta) {
    numerator <- x ^ beta
    denominator <- alpha ^ beta + x ^ beta
    return(numerator / denominator)
}

log_logistic_survivor <- function(x, alpha, beta) {
    return(1 - log_logistic_cdf(x, alpha, beta))
}
```

```r
log_logistic_hazard <- function(x, alpha, beta) {
    return(log_logistic_density(x, alpha, beta) / log_logistic_survivor(x, alpha, beta))
}
```

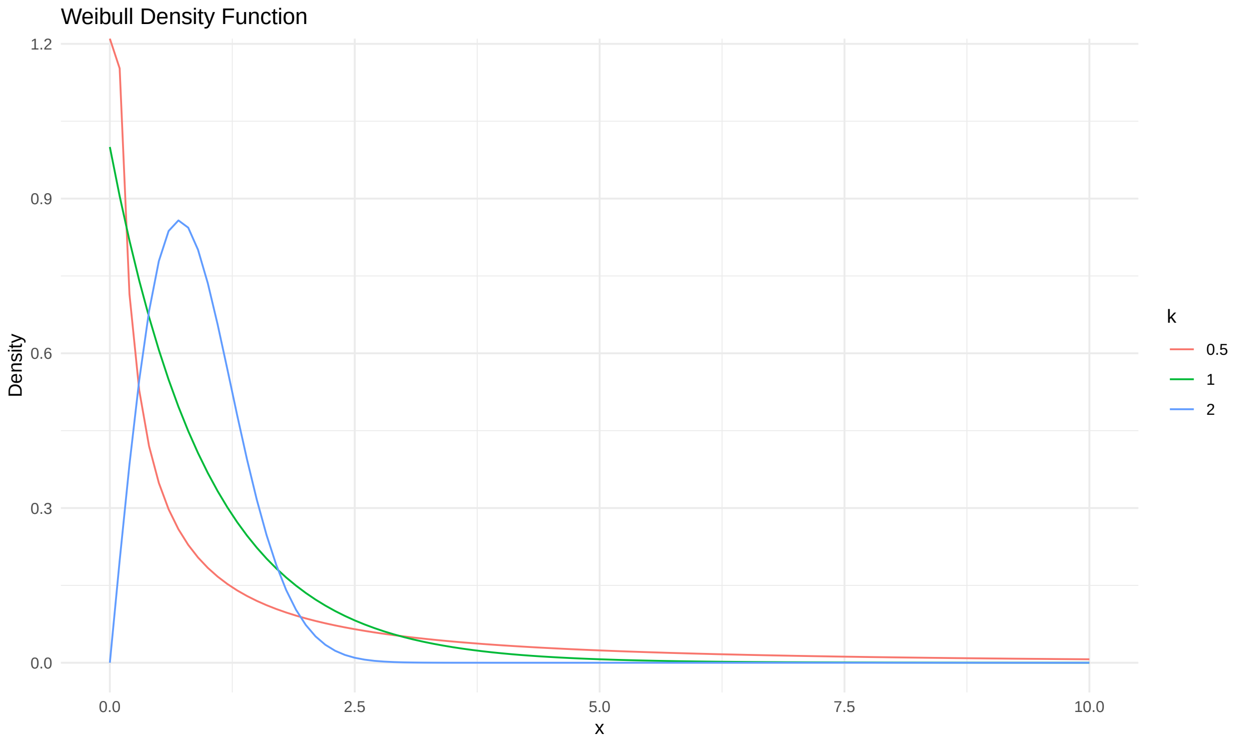
(a)

```r
alpha <- 1
beta <- c(0.25, 0.5, 1, 4, 8)
x <- seq(0, 10, 0.1)

# Density function
density_data <- expand.grid(x = x, beta = beta) %>%
  mutate(density = log_logistic_density(x, alpha, beta))

p1 <- ggplot(density_data, aes(x = x, y = density, color = as.factor(beta))) +
  geom_line() +
  labs(title = "Log-logistic Density Function", x = "x", y = "Density", color = "Beta") +
  theme_minimal()

# CDF
cdf_data <- expand.grid(x = x, beta = beta) %>%
  mutate(density = log_logistic_cdf(x, alpha, beta))

p2 <- ggplot(cdf_data, aes(x = x, y = density, color = as.factor(beta))) +
    geom_line() +
    labs(title = "Log-logistic CDF", x = "x", y = "CDF", color = "Beta") +
    theme_minimal()

# Survivor function
survivor_data <- expand.grid(x = x, beta = beta) %>%
  mutate(density = log_logistic_survivor(x, alpha, beta))

p3 <- ggplot(survivor_data, aes(x = x, y = density, color = as.factor(beta))) +
    geom_line() +
    labs(title = "Log-logistic Survivor Function", x = "x", y = "Survivor Function", color = "Beta") +
    theme_minimal()

# Hazard function
hazard_data <- expand.grid(x = x, beta = beta) %>%
  mutate(density = log_logistic_hazard(x, alpha, beta))

p4 <- ggplot(hazard_data, aes(x = x, y = density, color = as.factor(beta))) +
    geom_line() +
    labs(title = "Log-logistic Hazard Function", x = "x", y = "Hazard Function", color = "Beta") +
    theme_minimal()

grid.arrange(p1, p2, p3, p4, ncol = 2)
```



(b)

```r
sigma <- 1
k <- c(0.5, 1, 2)
x <- seq(0, 10, 0.1)

# k1
density_data <- expand.grid(x = x, k = k) %>%
  mutate(density = dweibull(x, k, sigma))

ggplot(density_data, aes(x = x, y = density, color = as.factor(k))) +
    geom_line() +
    labs(title = "Weibull Density Function", x = "x", y = "Density", color = "k") +
    theme_minimal()
```



Problem 2

(a)

```
pbc_data <- read.table("data/pbc.txt", header = TRUE)
pbc_data$treat <- factor(pbc_data$treat, levels = c(1, 2), labels = c("Placebo", "Treatment"))
skim(pbc_data)
```

Table 1: Data summary

Name	pbc_data
Number of rows	184
Number of columns	13
Column type frequency:	
factor	1
numeric	12
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
treat	0	1	FALSE	2	Tre: 94, Pla: 90

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
X.id.	0	1	224.76	241.86	5.00	99.75	154.50	220.25	961.00	
time	0	1	4.06	2.74	0.02	1.67	3.60	5.99	11.64	
d	0	1	0.52	0.50	0.00	0.00	1.00	1.00	1.00	
age	0	1	55.07	10.66	25.00	48.00	55.00	63.00	78.00	
logb0	0	1	0.01	0.45	-0.82	-0.35	-0.03	0.33	1.20	
alb0	0	1	-0.06	5.88	-14.32	-4.32	-0.32	3.68	22.18	
cenc0	0	1	0.17	0.38	0.00	0.00	0.00	0.00	1.00	
cir0	0	1	0.29	0.45	0.00	0.00	0.00	1.00	1.00	
X_st	0	1	1.00	0.00	1.00	1.00	1.00	1.00	1.00	
X_d	0	1	0.52	0.50	0.00	0.00	1.00	1.00	1.00	
X_t	0	1	4.06	2.74	0.02	1.67	3.60	5.99	11.64	
X_t0	0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

```
pbc_data$surv <- Surv(pbc_data$time, pbc_data$cenc0)
head(pbc_data$surv)
```

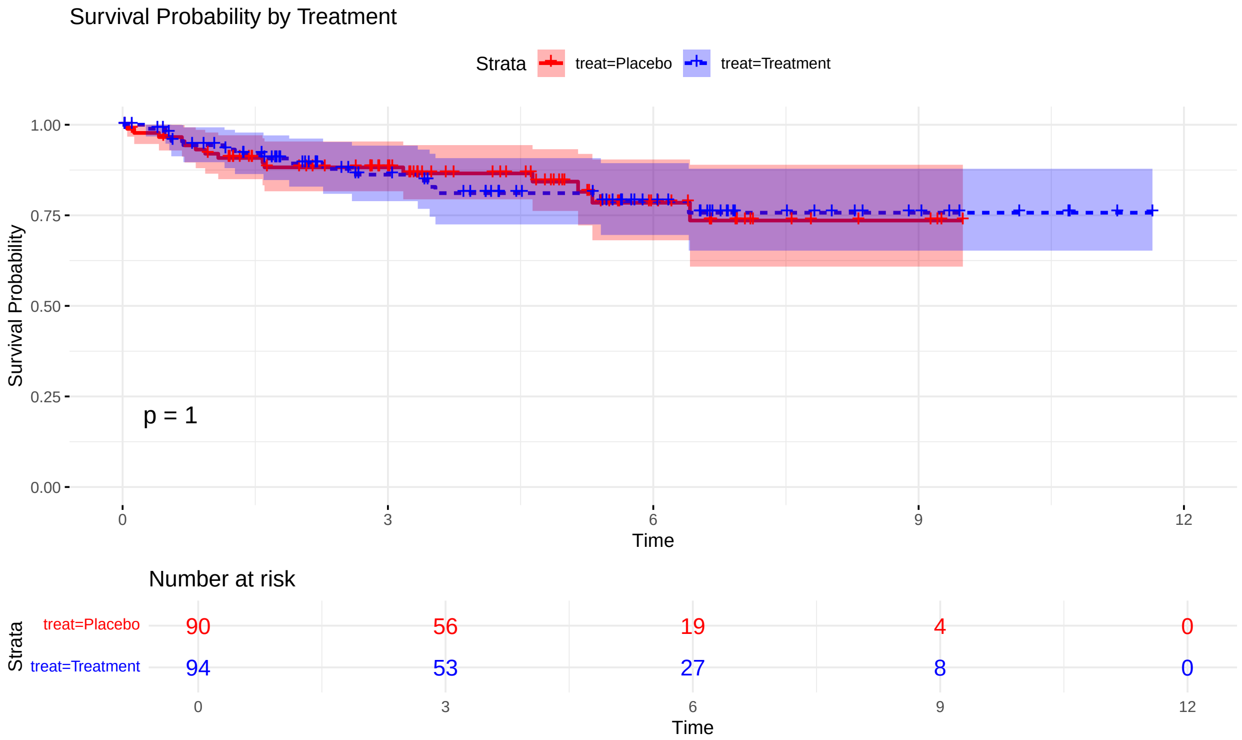
```
## [1] 1.582478  9.032169+ 2.286105+ 2.078029+ 5.311430  2.633812+
```

(b)

```
sfit <- survfit(surv ~ treat, data = pbc_data)
summary(sfit)
```

```
## Call: survfit(formula = surv ~ treat, data = pbc_data)
##
##                 treat=Placebo
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  0.052     89       1    0.989  0.0112        0.967        1.000
##  0.131     87       1    0.977  0.0158        0.947        1.000
##  0.411     86       1    0.966  0.0193        0.929        1.000
##  0.671     84       1    0.955  0.0222        0.912        0.999
##  0.701     83       1    0.943  0.0247        0.896        0.993
##  0.827     82       1    0.932  0.0270        0.880        0.986
##  0.934     81       1    0.920  0.0290        0.865        0.979
##  1.081     79       1    0.908  0.0309        0.850        0.971
##  1.582     70       1    0.895  0.0331        0.833        0.963
##  1.604     69       1    0.882  0.0350        0.816        0.954
##  3.173     53       1    0.866  0.0381        0.794        0.944
##  4.635     38       1    0.843  0.0434        0.762        0.932
##  5.150     30       1    0.815  0.0502        0.722        0.920
##  5.311     27       1    0.785  0.0567        0.681        0.904
##  6.415     16       1    0.736  0.0713        0.608        0.890
##
##                 treat=Treatment
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  0.263     91       1    0.989  0.0109        0.968        1.000
##  0.485     88       1    0.978  0.0155        0.948        1.000
##  0.528     86       1    0.966  0.0191        0.930        1.000
##  0.550     85       1    0.955  0.0220        0.913        0.999
##  0.687     83       1    0.944  0.0245        0.897        0.993
##  1.153     79       1    0.932  0.0270        0.880        0.986
##  1.270     77       1    0.919  0.0292        0.864        0.979
##  1.593     73       1    0.907  0.0314        0.847        0.971
##  1.884     66       1    0.893  0.0338        0.829        0.962
##  2.267     59       1    0.878  0.0365        0.809        0.952
##  2.593     56       1    0.862  0.0390        0.789        0.942
##  3.337     52       1    0.846  0.0417        0.768        0.931
##  3.472     49       1    0.828  0.0442        0.746        0.920
##  3.537     48       1    0.811  0.0466        0.725        0.908
##  5.407     36       1    0.789  0.0504        0.696        0.894
##  6.401     25       1    0.757  0.0574        0.653        0.879
```

```r
sfit_data <- as.data.frame(sfit$time)
sfit_data$surv <- sfit$surv
sfit_data$strata <- rep(names(sfit$strata), sfit$strata)

ggsurvplot(
  sfit,
  data = pbc_data,
  pval = TRUE,
  conf.int = TRUE,
  risk.table = TRUE,
  risk.table.col = "strata",
  linetype = "strata",
  palette = c("red", "blue"),
  xlab = "Time",
  ylab = "Survival Probability",
  title = "Survival Probability by Treatment",
  ggtheme = theme_minimal()
)
```



We cannot see a clear difference in survival between the two treatments. In fact, the Placebo group even has a higher Survival Probability for a while.

(c)

```r
sdiff <- survdiff(surv ~ treat, data = pbc_data)
sdiff
```

```
## Call:
## survdiff(formula = surv ~ treat, data = pbc_data)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## treat=Placebo   90       15       15  2.89e-06  5.62e-06
## treat=Treatment 94       16       16  2.70e-06  5.62e-06
##
##  Chisq= 0  on 1 degrees of freedom, p= 1
```

We get a p-value of 1, which means that there is no significant difference in survival between the two treatments.

(d)

```r
cox_model <- coxph(surv ~ treat, data = pbc_data)
summary(cox_model)
```

```
## Call:
## coxph(formula = surv ~ treat, data = pbc_data)
##
##   n= 184, number of events= 31
##
##                      coef  exp(coef)   se(coef)      z Pr(>|z|)
## treatTreatment -0.0008535  0.9991469  0.3601565 -0.002    0.998
##
##                exp(coef) exp(-coef) lower .95 upper .95
## treatTreatment    0.9991      1.001    0.4932     2.024
##
## Concordance= 0.499  (se = 0.048 )
## Likelihood ratio test= 0  on 1 df,   p=1
## Wald test            = 0  on 1 df,   p=1
## Score (logrank) test = 0  on 1 df,   p=1
```

We get a p-value of 1, which means that there is no significant difference in survival between the two treatments.

(e)

```
cox_model2 <- coxph(surv ~ treat + age + cenc0, data = pbc_data)
summary(cox_model2)
```

```
## Call:
## coxph(formula = surv ~ treat + age + cenc0, data = pbc_data)
##
##   n= 184, number of events= 31
##
##                      coef  exp(coef)   se(coef)      z Pr(>|z|)
## treatTreatment -5.802e-02  9.436e-01  3.794e-01 -0.153    0.878
## age             1.365e-02  1.014e+00  1.751e-02  0.779    0.436
## cenc0           2.319e+01  1.181e+10  6.201e+03  0.004    0.997
##
##                exp(coef) exp(-coef) lower .95 upper .95
## treatTreatment 9.436e-01  1.060e+00    0.4486     1.985
## age            1.014e+00  9.864e-01    0.9795     1.049
## cenc0          1.181e+10  8.468e-11    0.0000       Inf
##
## Concordance= 0.949  (se = 0.011 )
## Likelihood ratio test= 141.8  on 3 df,   p=<2e-16
## Wald test            = 0.61  on 3 df,   p=0.9
## Score (logrank) test = 253.8  on 3 df,   p=<2e-16
```

By adding both age and cenc0 as predictors, we get a highly significant p-value in both the Likelihood ratio and logrank test, but surprisingly, the p-value for the Wald test is 0.9. This could be due to the fact that we have a small sample size. We can also observe a high Concordance of 0.949, which indicates that the model is good at predicting the data.

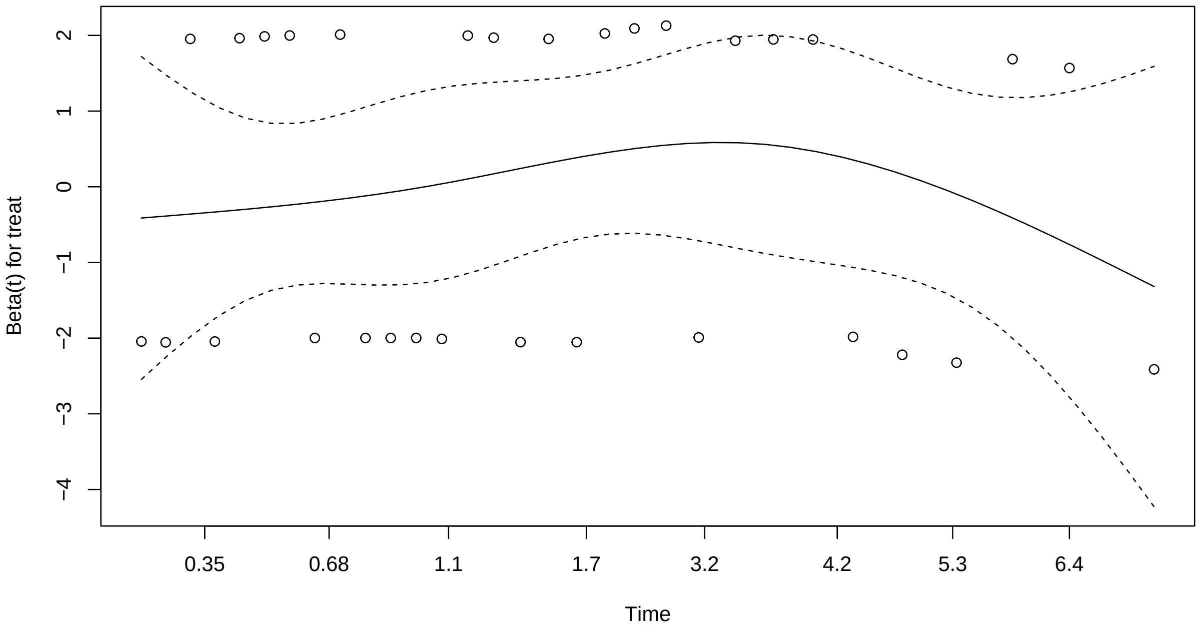
The hazard rate for utilizing the treatment is by a factor of 0.944 smaller than without the treatment, for age it is by a factor of 1.014 and for cenc0 it is 1.181.

(f)

```
# Model 1: Only treatment
cox.zph(cox_model)
```

```
##           chisq df    p
## treat  0.00459  1 0.95
## GLOBAL 0.00459  1 0.95
```

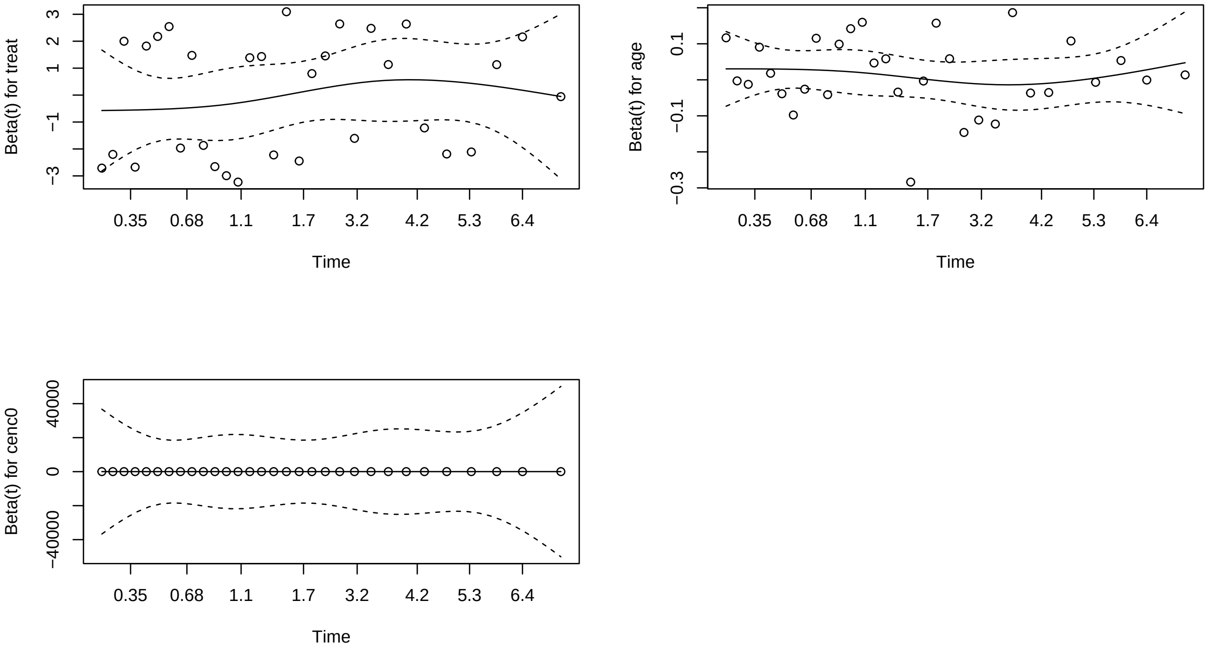
```
plot(cox.zph(cox_model))
```



```
# Model 2: Treatment, age and cenc0
par(mfrow = c(2, 2))
cox.zph(cox_model2)
```

```
##            chisq df    p
## treat  8.07e-01  1 0.37
## age    1.11e-01  1 0.74
## cenc0  6.13e-09  1 1.00
## GLOBAL 1.14e+00  3 0.77
```

```
plot(cox.zph(cox_model2))
```


Beta(t) for treat
3
2
1
−1
−3
0.35
0.68
1.1
1.7
3.2
4.2
5.3
6.4
Time
Beta(t) for age
0.1
−0.1
−0.3
0.35
0.68
1.1
1.7
3.2
4.2
5.3
6.4
Time
Beta(t) for cenc0
40000
0
−40000
0.35
0.68
1.1
1.7
3.2
4.2
5.3
6.4
Time