Day9 exercise solutions

Ali Movasati

Nov. 13th, 2024

```
# Set global code chunk options
knitr::opts_chunk$set(warning = FALSE)
```

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

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

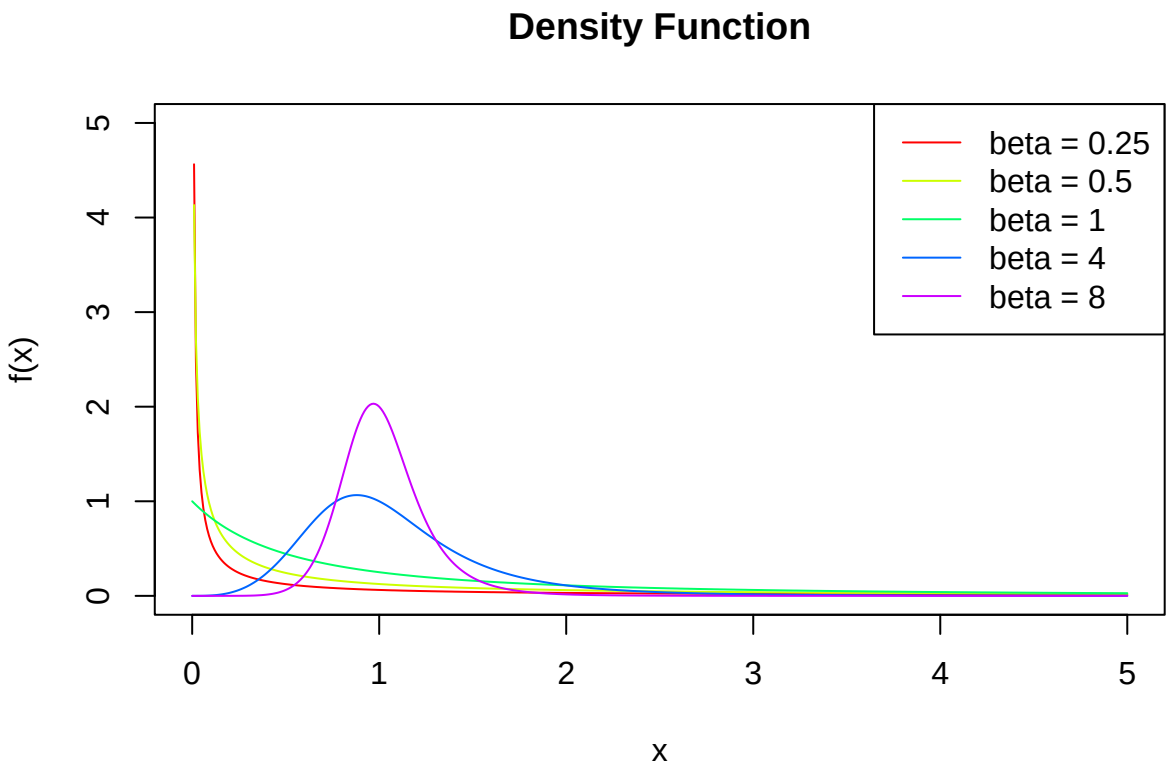
Problem 1

1.A)

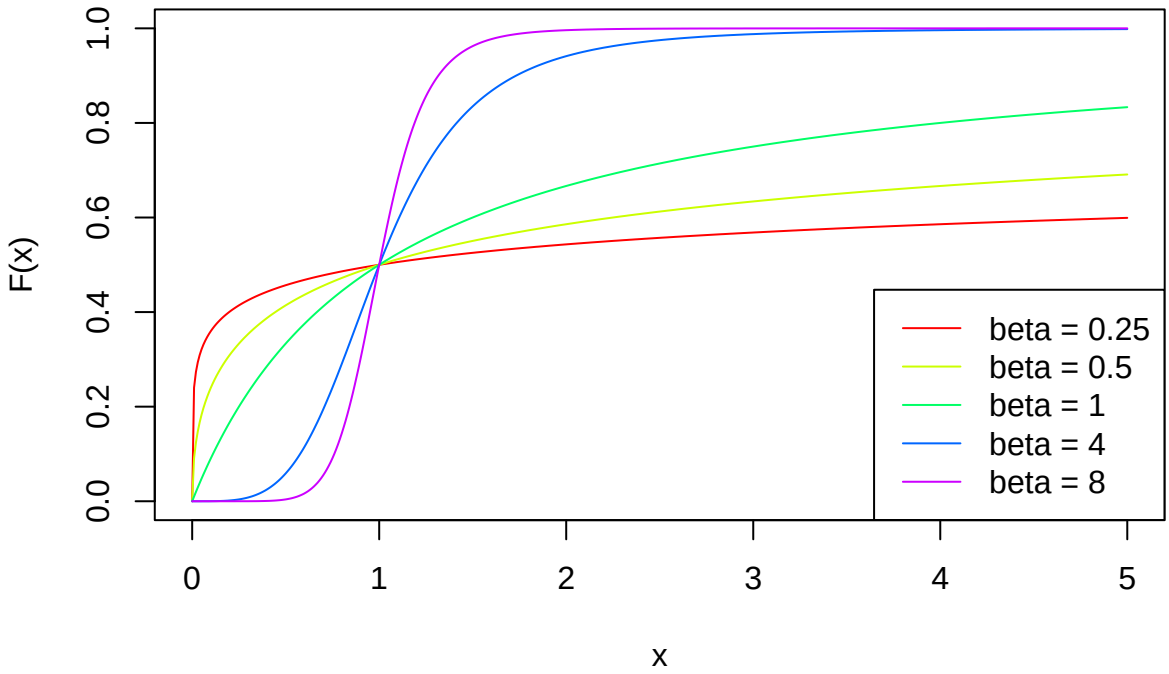
```
# Density function
log_logistic_density <- function(x, alpha, beta) {
  (beta/alpha) * (x/alpha)^(beta-1) / (1 + (x/alpha)^beta)^2
}

# Distribution function
log_logistic_dist <- function(x, alpha, beta) {
  (x^beta) / (alpha^beta + x^beta)
}

# Survival function
log_logistic_survival <- function(x, alpha, beta) {
  1 - log_logistic_dist(x, alpha, beta)
}

# Hazard function
log_logistic_hazard <- function(x, alpha, beta) {
  log_logistic_density(x, alpha, beta) / log_logistic_survival(x, alpha, beta)
}
```

```
# set parameters
x <- seq(0, 5, by = 0.01)
alpha <- 1
betas <- c(0.25, 0.5, 1, 4, 8)
```

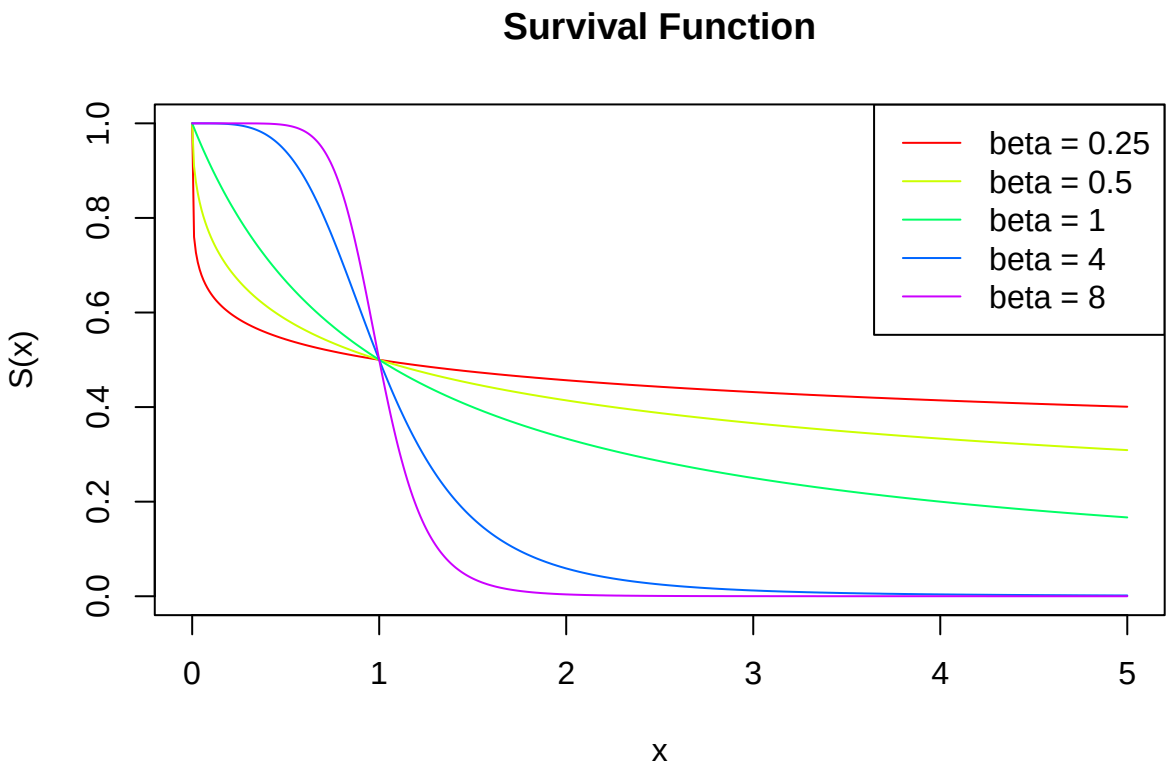
```r
colors <- rainbow(length(betas))


# Plot density
plot(NULL, xlim=c(0,5), ylim=c(0,5),
     main="Density Function", xlab="x", ylab="f(x)")
for(i in 1:length(betas)) {
  lines(x, log_logistic_density(x, alpha, betas[i]), col=colors[i])
}
legend("topright", legend=paste("beta =", betas), col=colors, lty=1)
```



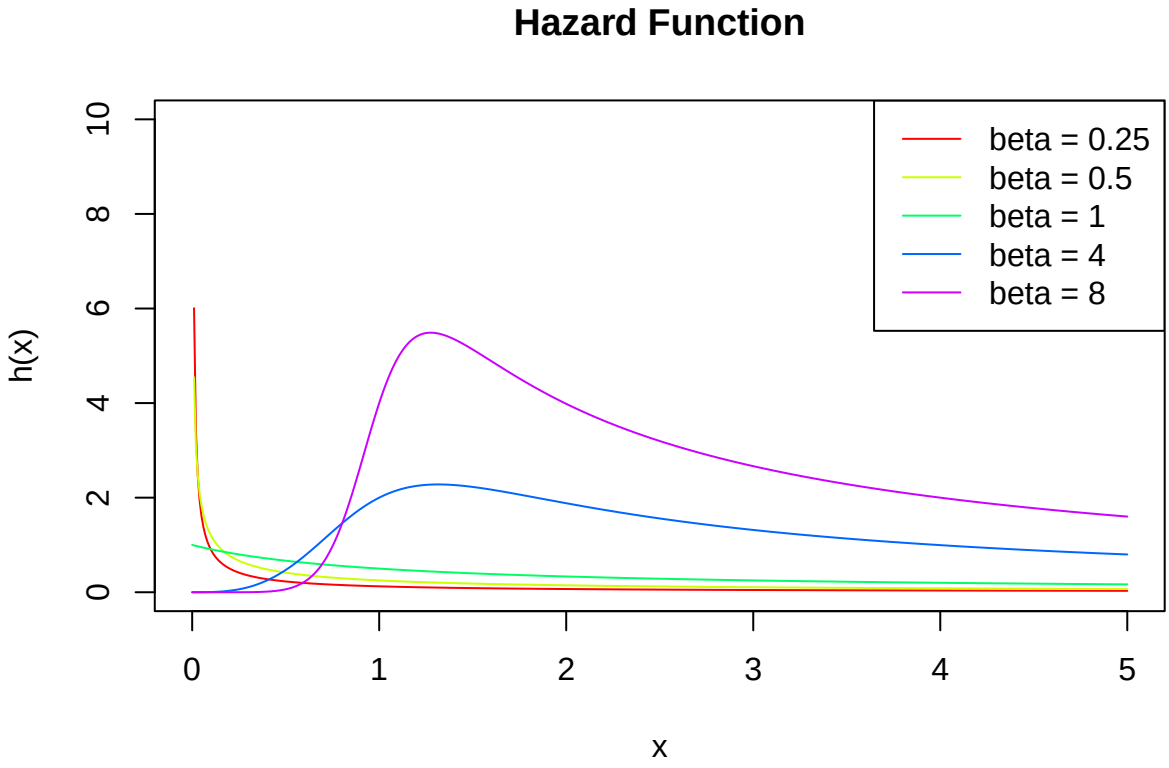

```r
# Plot distribution function
plot(NULL, xlim=c(0,5), ylim=c(0,1),
     main="Distribution Function", xlab="x", ylab="F(x)")
for(i in 1:length(betas)) {
  lines(x, log_logistic_dist(x, alpha, betas[i]), col=colors[i])
}
legend("bottomright", legend=paste("beta =", betas), col=colors, lty=1)
```




```r
# Plot survival function
plot(NULL, xlim=c(0,5), ylim=c(0,1),
     main="Survival Function", xlab="x", ylab="S(x)")
for(i in 1:length(betas)) {
  lines(x, log_logistic_survival(x, alpha, betas[i]), col=colors[i])
}
legend("topright", legend=paste("beta =", betas), col=colors, lty=1)
```



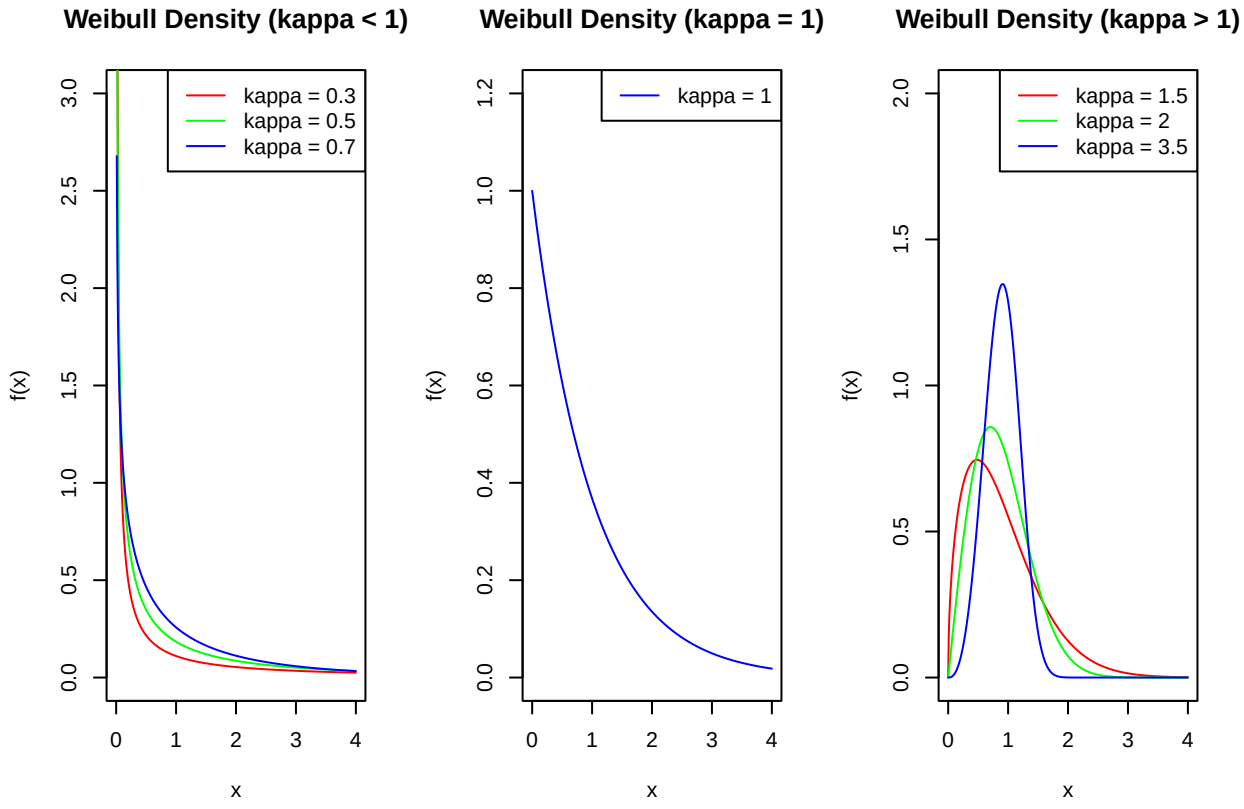

```r
# Plot hazard function
plot(NULL, xlim=c(0,5), ylim=c(0,10),
     main="Hazard Function", xlab="x", ylab="h(x)")
for(i in 1:length(betas)) {
  lines(x, log_logistic_hazard(x, alpha, betas[i]), col=colors[i])
}
legend("topright", legend=paste("beta =", betas), col=colors, lty=1)
```



1.B)

```r
# Create plots to show different shapes for   < 1,   = 1, and   > 1 with   = 1

# Set up the plotting parameters
x <- seq(0, 4, by = 0.01)
sigma <- 1

# Case 1:   < 1
kappa_less_1 <- c(0.3, 0.5, 0.7)
# Case 2:   = 1 (exponential case)
kappa_equal_1 <- 1
# Case 3:   > 1
kappa_more_1 <- c(1.5, 2, 3.5)

# Weibull density function
weibull_density <- function(x, kappa, sigma) {
  (kappa/sigma) * (x/sigma)^(kappa-1) * exp(-(x/sigma)^kappa)
}

# Create plot for all cases
par(mfrow=c(1,3))

# Plot for   < 1
plot(NULL, xlim=c(0,4), ylim=c(0,3),
     main="Weibull Density (kappa < 1)",
     xlab="x", ylab="f(x)")
for(k in kappa_less_1) {
  lines(x, weibull_density(x, k, sigma),
```

```
        col=rainbow(length(kappa_less_1))[which(kappa_less_1 == k)])
}
legend("topright", legend=paste("kappa =", kappa_less_1),
       col=rainbow(length(kappa_less_1)), lty=1)

# Plot for  = 1
plot(NULL, xlim=c(0,4), ylim=c(0,1.2),
     main="Weibull Density (kappa = 1)",
     xlab="x", ylab="f(x)")
lines(x, weibull_density(x, kappa_equal_1, sigma), col="blue")
legend("topright", legend="kappa = 1", col="blue", lty=1)

# Plot for  > 1
plot(NULL, xlim=c(0,4), ylim=c(0,2),
     main="Weibull Density (kappa > 1)",
     xlab="x", ylab="f(x)")
for(k in kappa_more_1) {
  lines(x, weibull_density(x, k, sigma),
        col=rainbow(length(kappa_more_1))[which(kappa_more_1 == k)])
}
legend("topright", legend=paste("kappa =", kappa_more_1),
       col=rainbow(length(kappa_more_1)), lty=1)
```



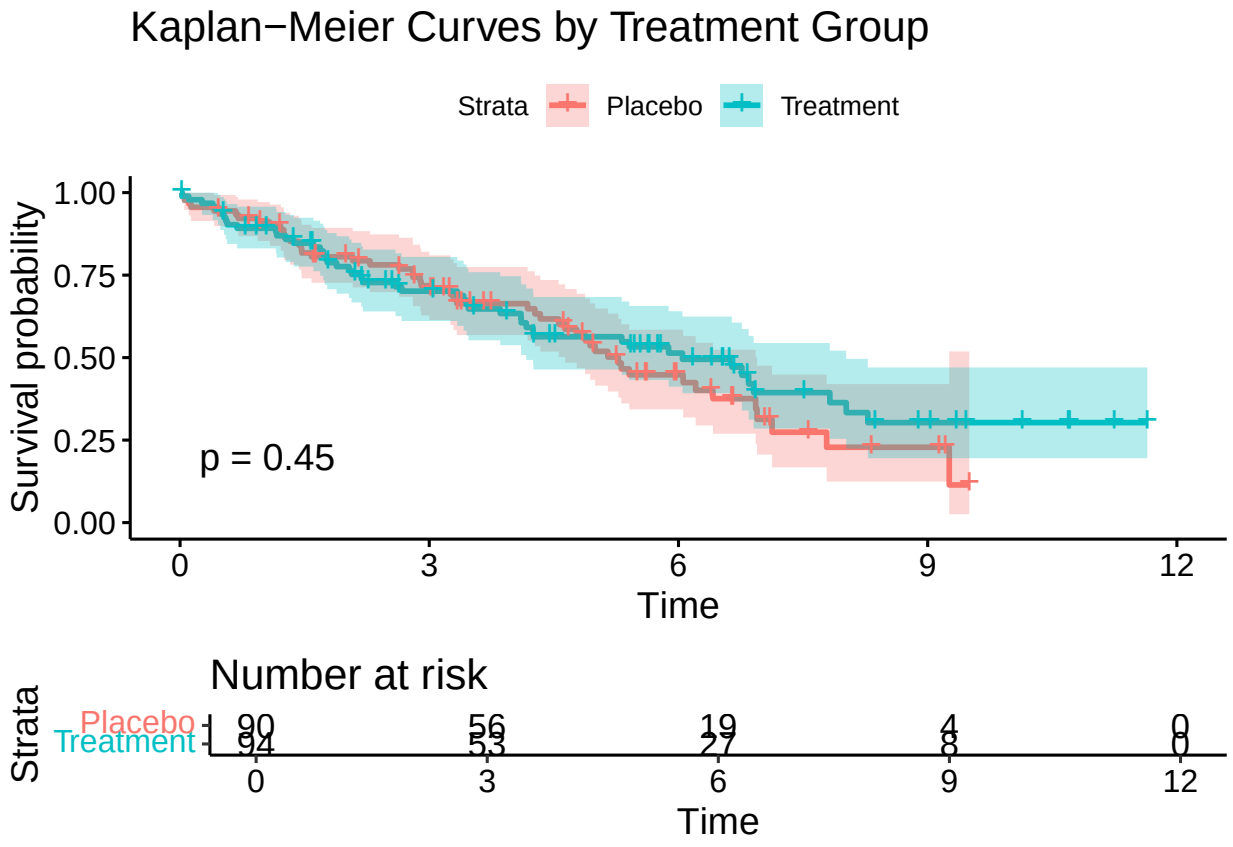
Problem 2

2.A)

```
# Load and prepare the data
pbc_data <- read.table("/Users/alimos313/Documents/studies/phd/university/courses/stat-modelling/StatMo

pbc_data$treat <- as.factor(pbc_data$treat)
surv_obj <- Surv(pbc_data$time, pbc_data$d)
```

2.B)

```
# Compare treatment groups using Kaplan-Meier curves
km_fit <- survfit(surv_obj ~ treat, data=pbc_data)

ggsurvplot(km_fit,
           data = pbc_data,
           pval = TRUE,
           conf.int = TRUE,
           risk.table = TRUE,
           title = "Kaplan-Meier Curves by Treatment Group",
           legend.labs = c("Placebo", "Treatment"))
```



<>

There appears to be improvement of the survival chance in the treated group compared to placebo group!
2.C)

```r
# Log-rank test

log_rank <- survdiff(surv_obj ~ treat, data=pbc_data)
print(log_rank)
```

```
## Call:
## survdiff(formula = surv_obj ~ treat, data = pbc_data)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## treat=1 90       49     45.3     0.297      0.57
## treat=2 94       47     50.7     0.265      0.57
##
##  Chisq= 0.6  on 1 degrees of freedom, p= 0.5
```

2.D)

```r
# Cox model with treatment only
cox_model1 <- coxph(surv_obj ~ treat, data=pbc_data)
summary(cox_model1)
```

```
## Call:
## coxph(formula = surv_obj ~ treat, data = pbc_data)
##
##   n= 184, number of events= 96
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## treat2 -0.1550    0.8564   0.2056 -0.754    0.451
##
##        exp(coef) exp(-coef) lower .95 upper .95
## treat2    0.8564      1.168    0.5724     1.281
##
## Concordance= 0.503  (se = 0.028 )
## Likelihood ratio test= 0.57  on 1 df,   p=0.5
## Wald test            = 0.57  on 1 df,   p=0.5
## Score (logrank) test = 0.57  on 1 df,   p=0.5
```

2.E)

```r
# Cox model with additional predictors
cox_model2 <- coxph(surv_obj ~ treat + age + cenc0, data=pbc_data)
summary(cox_model2)
```

```
## Call:
## coxph(formula = surv_obj ~ treat + age + cenc0, data = pbc_data)
##
##   n= 184, number of events= 96
##
##            coef exp(coef) se(coef)      z Pr(>|z|)
## treat2 -0.18754   0.82900  0.20592 -0.911 0.362430
## age     0.03504   1.03566  0.01017  3.445 0.000571 ***
## cenc0   1.34787   3.84921  0.24601  5.479 4.28e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
##        exp(coef) exp(-coef) lower .95 upper .95
## treat2     0.829     1.2063    0.5537     1.241
## age        1.036     0.9656    1.0152     1.057
## cenc0      3.849     0.2598    2.3766     6.234
##
## Concordance= 0.669  (se = 0.032 )
## Likelihood ratio test= 36.26  on 3 df,   p=7e-08
## Wald test            = 41.19  on 3 df,   p=6e-09
## Score (logrank) test = 45.65  on 3 df,   p=7e-10
```

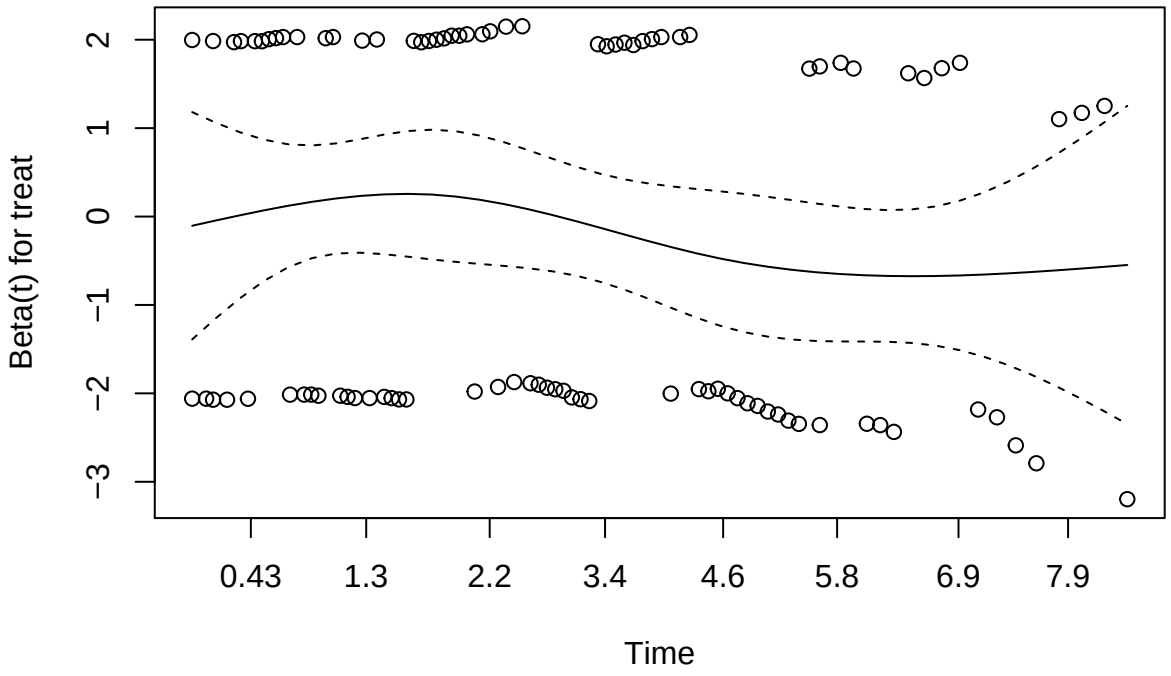
<>

2.F)

```
# Check proportional hazards assumption
# For model with treatment only
test.ph1 <- cox.zph(cox_model1)
print(test.ph1)
```

```
##        chisq df    p
## treat    2.1  1 0.15
## GLOBAL   2.1  1 0.15
```

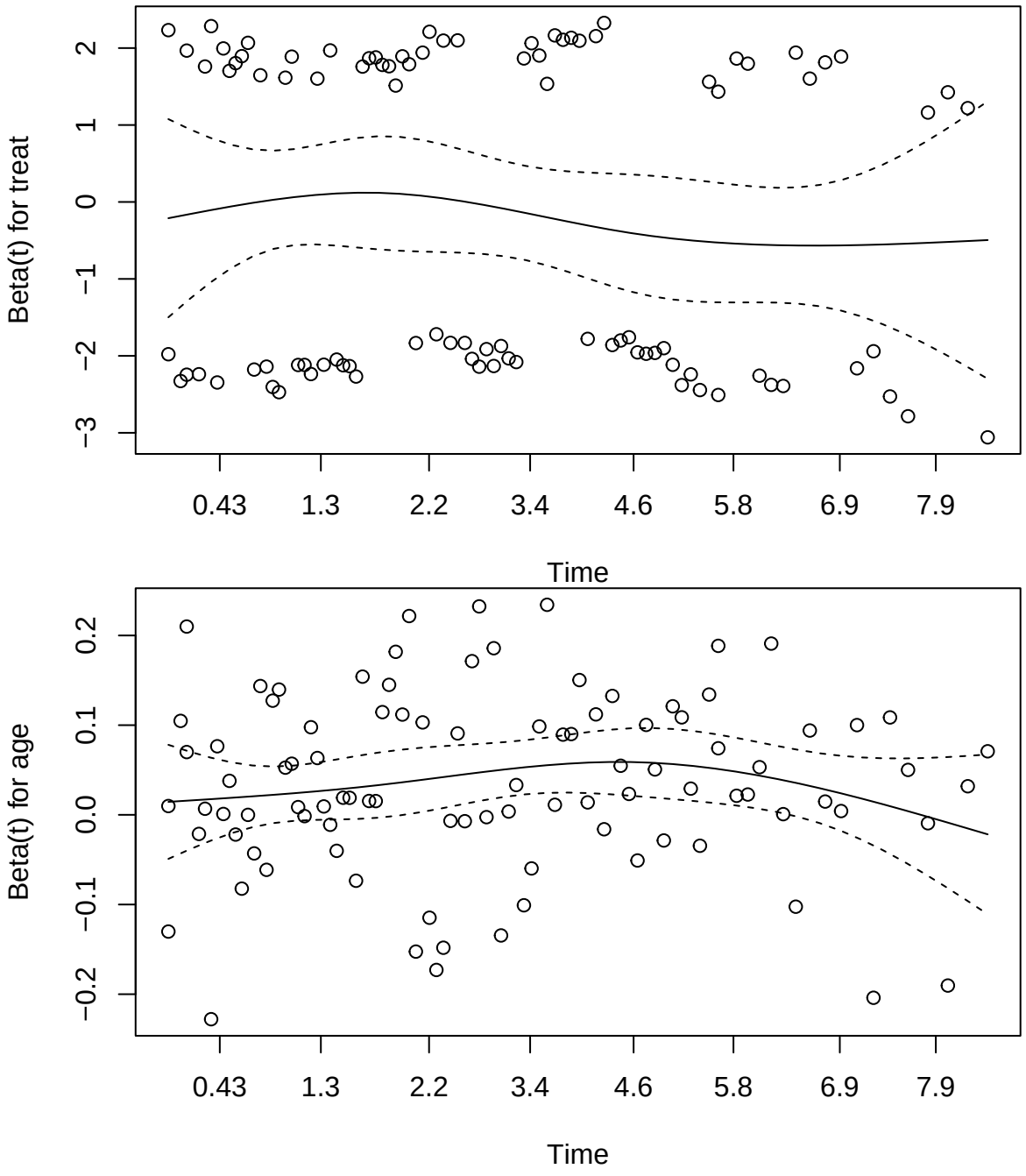
```
plot(test.ph1)
```

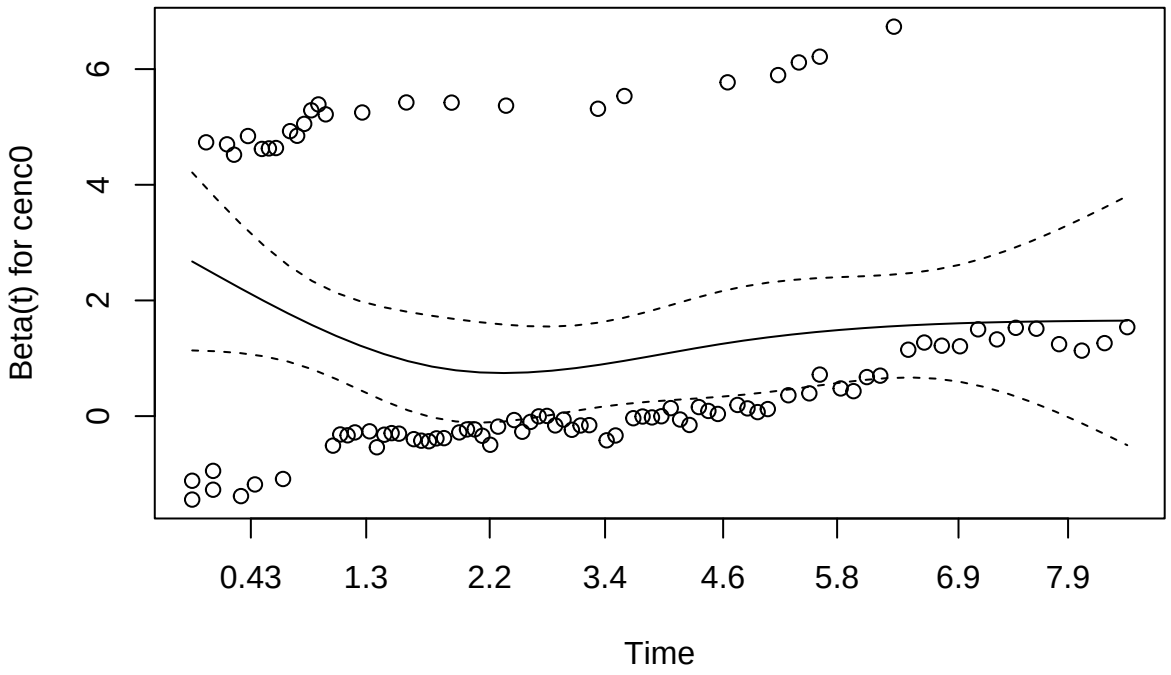



```
# For model with multiple predictors
test.ph2 <- cox.zph(cox_model2)
print(test.ph2)
```

```
##         chisq df    p
## treat  1.0383  1 0.31
## age    0.0411  1 0.84
## cenc0  0.2650  1 0.61
## GLOBAL 1.3281  3 0.72
```

```
plot(test.ph2)
```




Beta(t) for cenc0
0
2
4
6
0.43
1.3
2.2
3.4
4.6
5.8
6.9
7.9
Time