2. Perform Cox's proportional hazards regression

a. What is your conclusion regarding the central question of the study? Do the two treatment groups differ with respect to 90-day survival after adjustment for confounders?

```r
Lungca <- read.table("Lungca(1).txt",header=T,na.strings=c("."))
attach(Lungca)
library(survival)
### crude
fit.crude <- coxph(Surv(X9,Y)~X1, data=Lungca, method=c("breslow"))
summary(fit.crude)
```

```
## Call:
## coxph(formula = Surv(X9, Y) ~ X1, data = Lungca, method = c("breslow"))
##
##   n= 137, number of events= 75
##
##      coef exp(coef) se(coef)     z Pr(>|z|)
## X1 0.3870    1.4726   0.2338 1.655   0.0978 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##    exp(coef) exp(-coef) lower .95 upper .95
## X1     1.473     0.6791    0.9313     2.329
##
## Concordance= 0.54  (se = 0.03 )
## Rsquare= 0.02   (max possible= 0.993 )
## Likelihood ratio test= 2.78  on 1 df,   p=0.0957
## Wald test            = 2.74  on 1 df,   p=0.09783
## Score (logrank) test = 2.77  on 1 df,   p=0.09576
```

```r
### only covariates
fit.cov <- coxph(Surv(X9,Y)~X2+X3+X4+X5+X6+X7+X8, data=Lungca, method=c("breslow"))
summary(fit.cov)
```

```
## Call:
## coxph(formula = Surv(X9, Y) ~ X2 + X3 + X4 + X5 + X6 + X7 + X8,
##     data = Lungca, method = c("breslow"))
##
##   n= 137, number of events= 75
##
##          coef exp(coef)  se(coef)      z Pr(>|z|)
## X2  0.459485  1.583259  0.470161  0.977  0.32842
## X3  1.164289  3.203644  0.429391  2.711  0.00670 **
## X4  1.323377  3.756084  0.457157  2.895  0.00379 **
## X5 -0.047793  0.953331  0.006779 -7.050 1.79e-12 ***
## X6 -0.011394  0.988671  0.010179 -1.119  0.26298
## X7 -0.018997  0.981182  0.011531 -1.647  0.09947 .
## X8  0.549689  1.732714  0.288012  1.909  0.05632 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##    exp(coef) exp(-coef) lower .95 upper .95
## X2    1.5833     0.6316    0.6300    3.9788
## X3    3.2036     0.3121    1.3808    7.4326
## X4    3.7561     0.2662    1.5332    9.2017
## X5    0.9533     1.0490    0.9407    0.9661
## X6    0.9887     1.0115    0.9691    1.0086
## X7    0.9812     1.0192    0.9593    1.0036
## X8    1.7327     0.5771    0.9853    3.0471
##
## Concordance= 0.776  (se = 0.035 )
## Rsquare= 0.413   (max possible= 0.993 )
## Likelihood ratio test= 73.06  on 7 df,   p=3.547e-13
## Wald test            = 65.94  on 7 df,   p=9.742e-12
## Score (logrank) test = 75.63  on 7 df,   p=1.067e-13
```

```r
### adjusted
fit.adj <- coxph(Surv(X9,Y)~X1+X2+X3+X4+X5+X6+X7+X8, data=Lungca, method=c("breslow"))
summary(fit.adj)
```

```
## Call:
## coxph(formula = Surv(X9, Y) ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 +
##     X8, data = Lungca, method = c("breslow"))
##
##   n= 137, number of events= 75
##
##          coef exp(coef)  se(coef)      z Pr(>|z|)
## X1  0.445071  1.560601  0.254983  1.745  0.08090 .
## X2  0.478034  1.612900  0.471091  1.015  0.31023
## X3  1.263252  3.536905  0.433658  2.913  0.00358 **
## X4  1.229926  3.420975  0.460167  2.673  0.00752 **
## X5 -0.047738  0.953383  0.006755 -7.067 1.58e-12 ***
## X6 -0.014196  0.985905  0.010148 -1.399  0.16185
## X7 -0.021319  0.978907  0.011626 -1.834  0.06670 .
## X8  0.546963  1.727997  0.291005  1.880  0.06017 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##    exp(coef) exp(-coef) lower .95 upper .95
## X1    1.5606     0.6408    0.9468    2.5724
## X2    1.6129     0.6200    0.6406    4.0607
## X3    3.5369     0.2827    1.5118    8.2747
## X4    3.4210     0.2923    1.3882    8.4303
## X5    0.9534     1.0489    0.9408    0.9661
## X6    0.9859     1.0143    0.9665    1.0057
## X7    0.9789     1.0215    0.9569    1.0015
## X8    1.7280     0.5787    0.9769    3.0567
##
## Concordance= 0.781  (se = 0.035 )
## Rsquare= 0.426   (max possible= 0.993 )
## Likelihood ratio test= 76.12  on 8 df,   p=2.939e-13
## Wald test            = 68.41  on 8 df,   p=1.017e-11
## Score (logrank) test = 79.02  on 8 df,   p=7.716e-14
```

H0: all coefficients=0 or the set of variables is not associated with survival;
H1: at least one coefficients is not 0 or one variables in the set is associated with survival.

Multivariate Model LR chi-square = 76.12 on 8 df, p=2.939e-13. Since the p-value is less than 0.05, with 95% confidence we reject the H0, and conclude that at least one variables in the set is associated with survival(chi-square = 76.12, p=2.939e-13).

We can also get the below estimated coefficients for each variable, and their p-values.

```
 coef exp(coef)  se(coef)      z Pr(>|z|)
```

X1 0.445071 1.560601 0.254983 1.745 0.08090
X2 0.478034 1.612900 0.471091 1.015 0.31023
X3 1.263252 3.536905 0.433658 2.913 0.00358
X4 1.229926 3.420975 0.460167 2.673 0.00752

X5 -0.047738 0.953383 0.006755 -7.067 1.58e-12
X6 -0.014196 0.985905 0.010148 -1.399 0.16185
X7 -0.021319 0.978907 0.011626 -1.834 0.06670
X8 0.546963 1.727997 0.291005 1.880 0.06017

Here we want to test for treatments (X1). H0: coefficient for treatment is 0; H1: coefficient for treatment not equal to 0. LR test: LR chi-square = Full Model chi-square - Model without X1 chi-sauare = 76.12-73.06 = 3.06 with 1 df. p-value = 0.08 $>$ 0.05. With 95% confidence we can not reject the H0, and conclude that coefficient for X1 is 0, so the two treatment groups do not differ with respect to 90-day survival after adjustment for confounders(chi-square = 3.06, p-value = 0.08).

b. What is the estimated risk of dying in the test group compared to the standard group? Include a 95% confidence interval. Show your work. Make a statement regarding the effect of treatment on survival with these results.

The estimated coefficient for treatment is 0.4451, so Hazard Ratio = exp(coefficient) = exp(0.4451) = 1.56. After adjustment for confounders, the relative risk of dying in the test group compared to the standard group is 1.56.

95% confidence interval: (0.95, 2.57). With 95% confidence, we estimate that the true hazard ratio for dying lies between 0.95 and 2.57. This confidence interval includes 1. There is not statistically significant evidence to reject H0 and suggest an increased hazard ratio of dying in the test group after adjustment for confounders.

c. Is there confounding in examining this association? Provide evidence.

The adjusted analysis provides hazard ratio is 1.56 with 95% CI (0.95, 2.57) and p-value is 0.0809. The hazard ratio from crude analysis is 1.47 with 95% CI (0.93, 2.33) and p-value is 0.0978. The two estimates differ by less than 10%, suggesting that there is no confounding in examining this association.

d. What other variables are associated with 90-day survival? Provide evidence.

H0: all coefficients=0 or the set of variables is not associated with survival;
H1: at least one coefficients is not 0 or one variables in the set is associated with survival.

From the output,
coef exp(coef) se(coef) z Pr(>|z|)
X1 0.445071 1.560601 0.254983 1.745 0.08090
X2 0.478034 1.612900 0.471091 1.015 0.31023

X3 1.263252 3.536905 0.433658 2.913 0.00358
X4 1.229926 3.420975 0.460167 2.673 0.00752

X5 -0.047738 0.953383 0.006755 -7.067 1.58e-12

X6 -0.014196 0.985905 0.010148 -1.399 0.16185
X7 -0.021319 0.978907 0.011626 -1.834 0.06670
X8 0.546963 1.727997 0.291005 1.880 0.06017

Cell type X3, X4 and Performance status X5 are associated with 90-day survival, since their p-values are less than 0.05, which is 0.004, 0.007 and 1.58e-12, respectively.

Cell type squamous,: HR.X3=3.54 (CI: 1.51, 8.27), Cell type small: HR.X4=3.42 (CI: 1.39, 8.43), Performance status: HR.X5=0.95 (CI: 0.94, 0.97). These confidence intervals also exclude 1. There is statistically significant evidence to reject H0 and suggest an increased hazard ratio of dying in the cell type squamous, small and performance status after adjustment for confounders.

After adjustment for confounders, the relative risk of dying in the squamous cell type compared to those without squamous cell type is 3.54 After adjustment for confounders, the relative risk of dying in the small cell type compared to those without small cell type is 3.42. After adjustment for confounders, for every additional value of performance status, the risk of dying falls by 5% (hazard ratio 0.95).

Since X3 and X4 are both type, we also perform a likelihood ratio test of the 3 variables associated with cell type: X2, X3 and X4. H0: type is not associated with 90-day survival after adjustment for confounders; H1: type is associated with 90-day survival after adjustment for confounders.

```
fit1 <- coxph(Surv(X9,Y)~X1 + X5 + X6 + X7 + X8, data=Lungca, method=c("breslow"))
summary(fit1)
```

```
## Call:
## coxph(formula = Surv(X9, Y) ~ X1 + X5 + X6 + X7 + X8, data = Lungca,
##     method = c("breslow"))
##
##   n= 137, number of events= 75
##
##          coef exp(coef)  se(coef)      z Pr(>|z|)
## X1  0.390698  1.478012  0.238441  1.639    0.101
## X5 -0.048958  0.952221  0.006519 -7.510 5.94e-14 ***
## X6 -0.009689  0.990357  0.010093 -0.960    0.337
## X7 -0.012331  0.987745  0.010712 -1.151    0.250
## X8  0.353689  1.424312  0.279484  1.266    0.206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##    exp(coef) exp(-coef) lower .95 upper .95
## X1    1.4780     0.6766    0.9262    2.3585
## X5    0.9522     1.0502    0.9401    0.9645
## X6    0.9904     1.0097    0.9710    1.0101
## X7    0.9877     1.0124    0.9672    1.0087
## X8    1.4243     0.7021    0.8236    2.4632
##
## Concordance= 0.752  (se = 0.035 )
## Rsquare= 0.362   (max possible= 0.993 )
## Likelihood ratio test= 61.51  on 5 df,   p=5.931e-12
## Wald test            = 60.45  on 5 df,   p=9.815e-12
## Score (logrank) test = 67.96  on 5 df,   p=2.717e-13
```

```r
fit2 <- coxph(Surv(X9,Y)~X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8, data=Lungca, method=c("breslow"))
summary(fit2)
```

```
## Call:
## coxph(formula = Surv(X9, Y) ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 +
##     X8, data = Lungca, method = c("breslow"))
##
##   n= 137, number of events= 75
##
##         coef exp(coef)  se(coef)      z Pr(>|z|)
## X1  0.445071  1.560601  0.254983  1.745  0.08090 .
## X2  0.478034  1.612900  0.471091  1.015  0.31023
## X3  1.263252  3.536905  0.433658  2.913  0.00358 **
## X4  1.229926  3.420975  0.460167  2.673  0.00752 **
## X5 -0.047738  0.953383  0.006755 -7.067 1.58e-12 ***
## X6 -0.014196  0.985905  0.010148 -1.399  0.16185
## X7 -0.021319  0.978907  0.011626 -1.834  0.06670 .
## X8  0.546963  1.727997  0.291005  1.880  0.06017 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##    exp(coef) exp(-coef) lower .95 upper .95
## X1    1.5606     0.6408    0.9468    2.5724
## X2    1.6129     0.6200    0.6406    4.0607
## X3    3.5369     0.2827    1.5118    8.2747
## X4    3.4210     0.2923    1.3882    8.4303
## X5    0.9534     1.0489    0.9408    0.9661
## X6    0.9859     1.0143    0.9665    1.0057
## X7    0.9789     1.0215    0.9569    1.0015
## X8    1.7280     0.5787    0.9769    3.0567
##
## Concordance= 0.781  (se = 0.035 )
## Rsquare= 0.426   (max possible= 0.993 )
## Likelihood ratio test= 76.12  on 8 df,   p=2.939e-13
## Wald test            = 68.41  on 8 df,   p=1.017e-11
## Score (logrank) test = 79.02  on 8 df,   p=7.716e-14
```

χ^2 = -2lnR = 76.12 - 61.51 = 14.61 with 3 df, p-value = 0.0001 Since the p-value is <0.05, we can reject the H0, and conclude that type is significantly associated with 90-day survival after adjustment for confounders(chi-square=14.61, p-value=0.0001).

In conclusion, type and performance status are also associated with 90-day survival after adjustment for confounder.

e. How do these results compare with those of the logistic regression analyses? Justify your responses.

	X1	X2	X3	X4	X5	X6	X7	X8
HR	1.561	1.613	3.537	3.421	0.953	0.986	0.979	1.728
OR	2.994	1.797	8.733	13.304	0.924	0.979	1.007	2.270

For most variables, Cox's proportional hazards regression gives the similar results as logistic regression, but for variables with high HR, logistic regression tends to give a higher OR.

Among the significant variables in the d part, the OR and HR for performance status(X5) is similar, but ORs for both types squamous(X3) and small(X4) are much higher than their HRs. From Cox Proportional Hazards Regression, Cell type squamous: HR.X3=3.54 (CI: 1.51, 8.27), Cell type small: HR.X4=3.42 (CI: 1.39, 8.43), Performance status: HR.X5=0.95 (CI: 0.94, 0.97).

For the logistic regression analyses results in homework 5, Cell type squamous: OR.X3=8.7333 (CI: 2.1897, 34.8317), Cell type small: OR.X4=13.3037 (CI: 2.5936, 68.2397), Performance status: OR.X5=0.9237 (CI: 0.8958, 0.95245).

From Cox's proportional hazards regression, we conclude that type and performance status are associated with 90-day survival after adjustment for confounders because the confidence intervals for HRs exclude 1. The confidence intervals for ORs we get from logistic regression analyses also exclude 1, so the conclusions of relation are the same, that is, type and performance status relate to 90-day survival significantly. However, since the confidence interval for HR of treatment (0.95, 2.57) includes 1, only logistic regression shows significant association between treatment and 90-day survival.

f. Is the proportional hazards assumption satisfied for the treatment groups? Is it satisfied for the other covariates? Explain your answers.

The proportional hazards assumption are, a. The natural logarithms of h(t|X) plotted against time are parallel for varying values of X, separated by beta. From the plot of log(t) and log(-log(S(t)), we can see that the two line about X1=1 and X1=0 are not parallel, so it does not suggest proportional hazards regression for the assumption of parallelity.

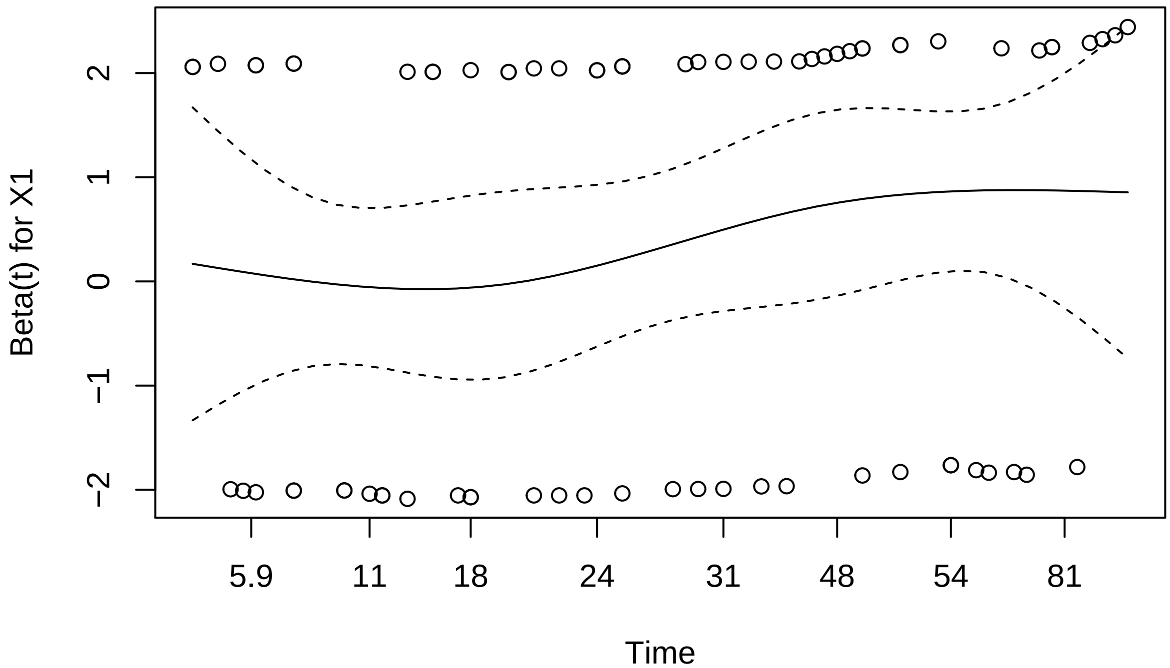
b. The Hazard Ratio does not depend on time. We test the proportionality of hazard with the function. $\beta(t) = \beta + \theta g(t)$ H0: $\theta = 0$ => proportional hazard is OK;
   Ha: $\theta \neq 0$ => proportional hazard is not satisfied.

For the crude analysis,

```
test <- cox.zph(fit.crude)
print(test)
```

```
##      rho chisq     p
## X1 0.169  2.16 0.142
```

```r
plot(test)
```



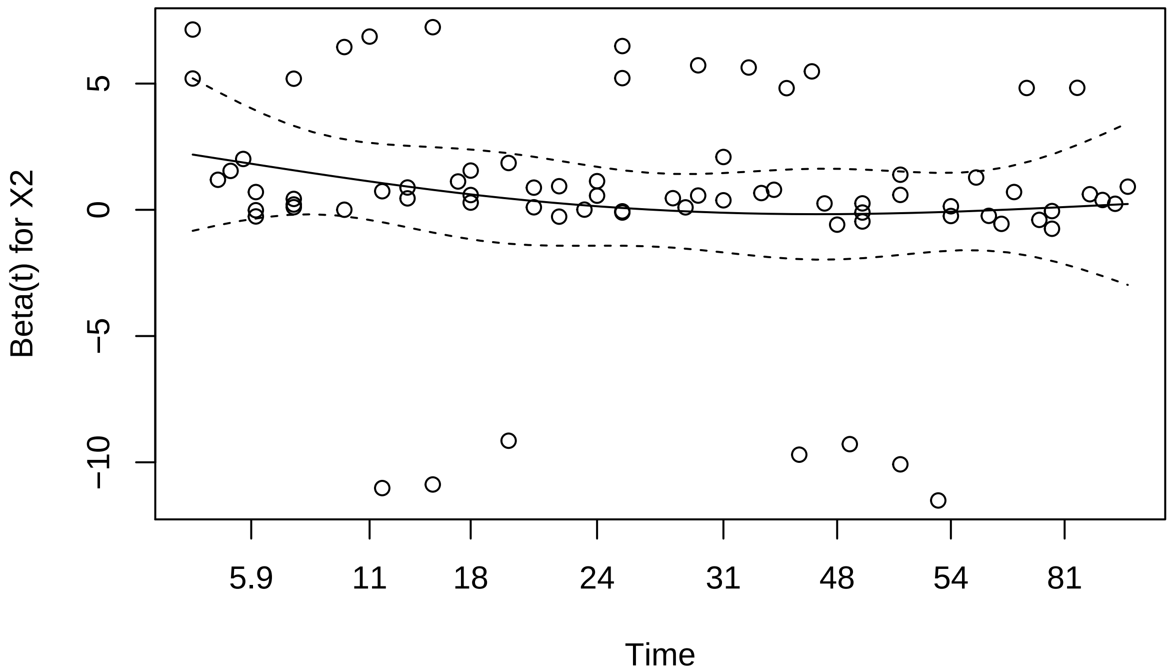
From the output, correlation is 0.169. Conclusions: We fail to reject the null hypothesis since the chi-square is 2.16 with a p-value 0.142 > 0.05, so this proportional hazards assumption is satisfied for the treatment groups.

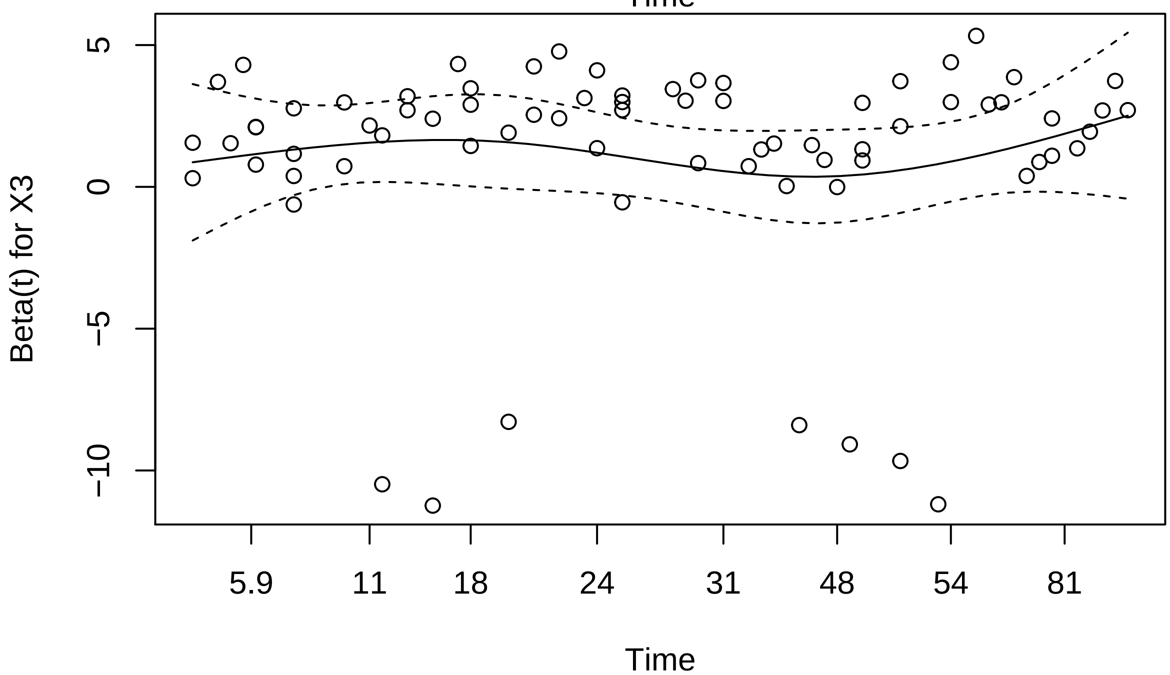
Only confounders:

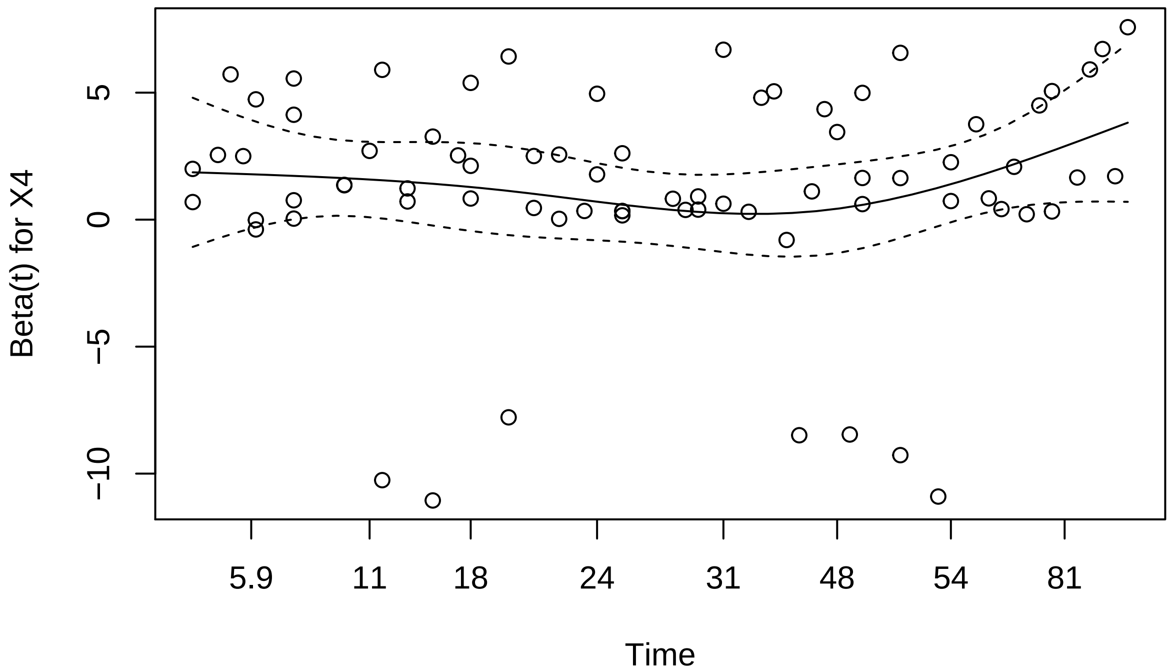
```r
test2 <- cox.zph(fit.cov)
print(test2)
```

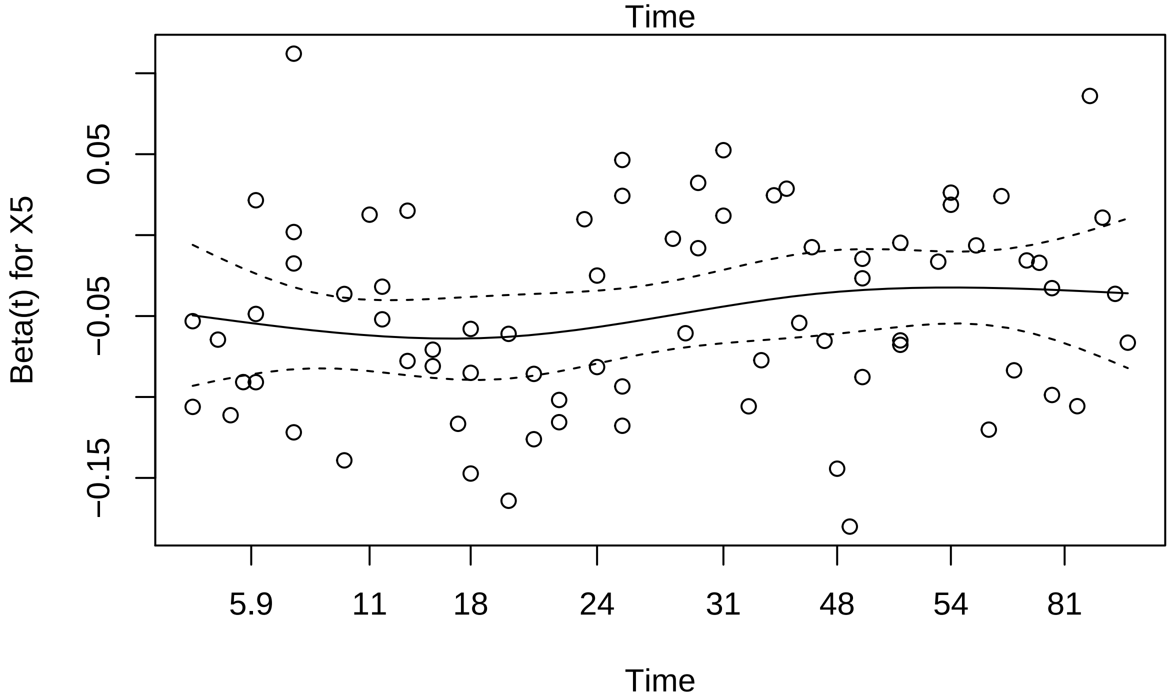
```
##             rho    chisq      p
## X2     -0.14481  1.56384 0.2111
## X3     -0.00453  0.00156 0.9685
## X4      0.03469  0.09450 0.7585
## X5      0.16823  2.20334 0.1377
## X6     -0.03097  0.07001 0.7913
## X7      0.23380  4.56045 0.0327
## X8     -0.03719  0.10279 0.7485
## GLOBAL       NA 11.19389 0.1304
```

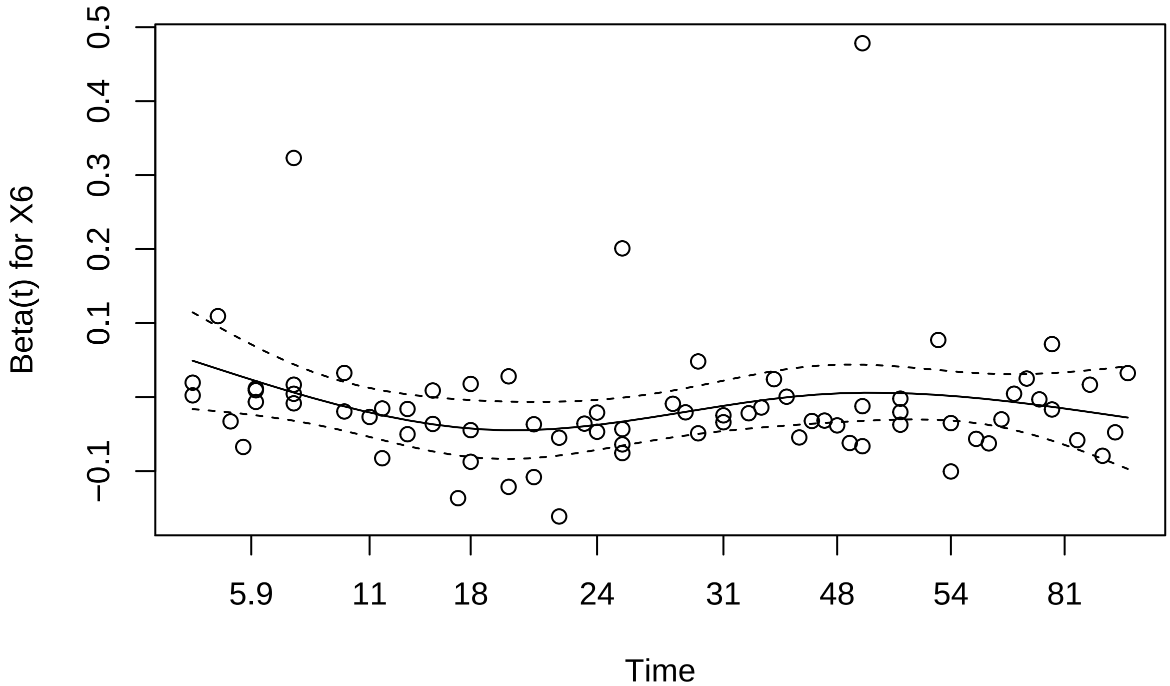
```r
plot(test2)
```

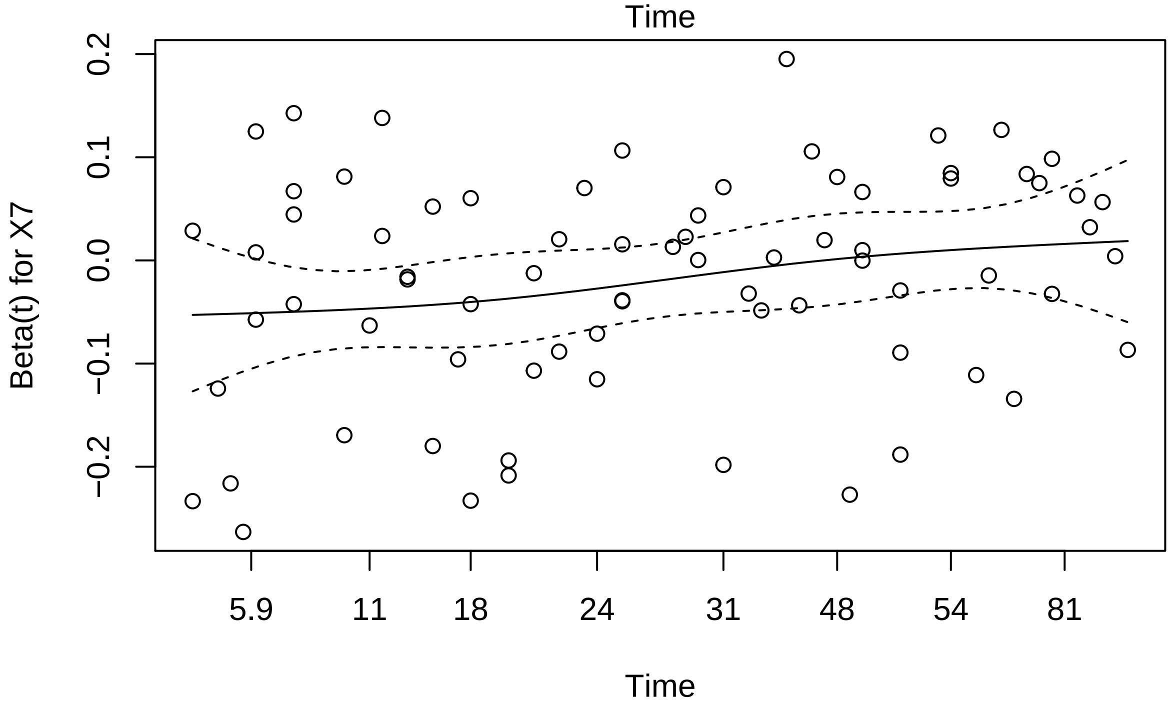

Beta(t) for X2
5
0
−5
−10
5.9
11
18
24
31
48
54
81
Time

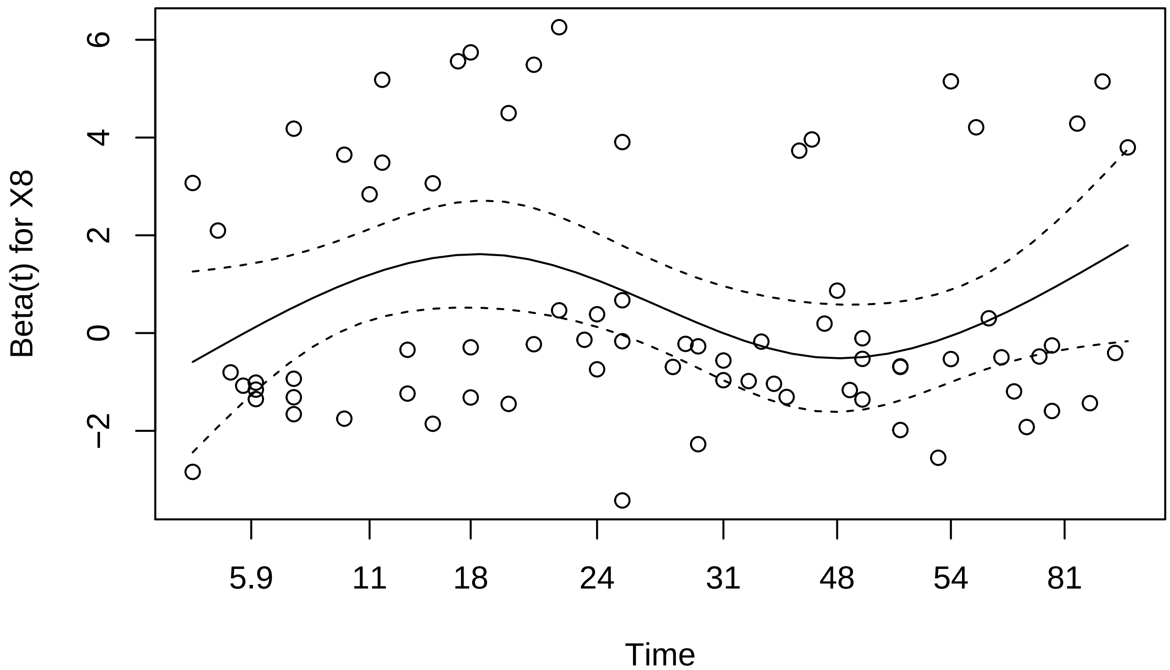

Beta(t) for X3
5
0
−5
−10
5.9
11
18
24
31
48
54
81
Time


Beta(t) for X4
5
0
−5
−10
5.9
11
18
24
31
48
54
81
Time


Beta(t) for X5
0.05
−0.05
−0.15
5.9
11
18
24
31
48
54
81
Time


Beta(t) for X6
0.5
0.4
0.3
0.2
0.1
−0.1
5.9
11
18
24
31
48
54
81
Time


Beta(t) for X7
0.2
0.1
0.0
−0.1
−0.2
5.9
11
18
24
31
48
54
81
Time



Conclusions: We fail to reject the null hypothesis since the p-value for variables X2, X3, X4, X5, X6 and X8 are larger than 0.05, so the proportional hazards assumption is satisfied for type (X2, X3, X4), performance status (X5), disease duration (X6) and therapy (X8). However, for the variable age (X7), this proportional hazards assumption is not satisfied since it has a small p-value of 0.0327.

3. These individuals were in reality followed much longer than 90 days, about 3 years. The real followup data are given in the last two fields on the data set: X10 is the survival time and X11 is the status at the end of the study. Repeat the proportional hazards analysis on the real survival time and evaluate whether treatment had an effect. What are your conclusions? Are the assumptions of the proportional hazards model met?

```
### crude
fit.crude_fo <- coxph(Surv(X10,X11)~X1, data=Lungca, method=c("breslow"))
summary(fit.crude_fo)
```

```
## Call:
## coxph(formula = Surv(X10, X11) ~ X1, data = Lungca, method = c("breslow"))
##
##   n= 137, number of events= 128
##
##       coef exp(coef) se(coef)    z Pr(>|z|)
## X1 0.01633   1.01646  0.18065 0.09    0.928
##
##    exp(coef) exp(-coef) lower .95 upper .95
## X1     1.016     0.9838    0.7134     1.448
##
## Concordance= 0.525  (se = 0.026 )
## Rsquare= 0   (max possible= 0.999 )
## Likelihood ratio test= 0.01  on 1 df,   p=0.928
```

```
## Wald test            = 0.01  on 1 df,   p=0.928
## Score (logrank) test = 0.01  on 1 df,   p=0.928
```

```r
### only covariates
fit.cov_fo <- coxph(Surv(X10,X11)~X2+X3+X4+X5+X6+X7+X8, data=Lungca, method=c("breslow"))
summary(fit.cov_fo)
```

```
## Call:
## coxph(formula = Surv(X10, X11) ~ X2 + X3 + X4 + X5 + X6 + X7 + 
##     X8, data = Lungca, method = c("breslow"))
## 
##   n= 137, number of events= 128 
## 
##         coef exp(coef)  se(coef)      z Pr(>|z|)    
## X2 -0.323679  0.723482  0.276822 -1.169   0.2423    
## X3  0.406885  1.502131  0.263835  1.542   0.1230    
## X4  0.867338  2.380564  0.297937  2.911   0.0036 ** 
## X5 -0.031960  0.968546  0.005506 -5.804 6.47e-09 ***
## X6  0.001215  1.001216  0.009289  0.131   0.8959    
## X7 -0.005759  0.994257  0.009128 -0.631   0.5281    
## X8  0.088170  1.092174  0.230533  0.382   0.7021    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
##    exp(coef) exp(-coef) lower .95 upper .95
## X2    0.7235     1.3822    0.4205    1.2447
## X3    1.5021     0.6657    0.8956    2.5193
## X4    2.3806     0.4201    1.3276    4.2686
## X5    0.9685     1.0325    0.9581    0.9791
## X6    1.0012     0.9988    0.9832    1.0196
## X7    0.9943     1.0058    0.9766    1.0122
## X8    1.0922     0.9156    0.6951    1.7160
## 
## Concordance= 0.735  (se = 0.03 )
## Rsquare= 0.352   (max possible= 0.999 )
## Likelihood ratio test= 59.45  on 7 df,   p=1.947e-10
## Wald test            = 59.71  on 7 df,   p=1.724e-10
## Score (logrank) test = 63.38  on 7 df,   p=3.185e-11
```

```r
### adjusted
fit.adj_fo <- coxph(Surv(X10,X11)~X1+X2+X3+X4+X5+X6+X7+X8, data=Lungca, method=c("breslow"))
summary(fit.adj_fo)
```

```
## Call:
## coxph(formula = Surv(X10, X11) ~ X1 + X2 + X3 + X4 + X5 + X6 + 
##     X7 + X8, data = Lungca, method = c("breslow"))
## 
##   n= 137, number of events= 128 
## 
##         coef exp(coef)  se(coef)      z Pr(>|z|)    
## X1  0.289936  1.336342  0.207210  1.399  0.16174    
## X2 -0.399628  0.670570  0.282663 -1.414  0.15742    
## X3  0.456859  1.579106  0.266273  1.716  0.08621 .  
```

```
## X4  0.788672  2.200471  0.302668  2.606  0.00917 **
## X5 -0.032622  0.967905  0.005505 -5.926 3.11e-09 ***
## X6 -0.000092  0.999908  0.009125 -0.010  0.99196
## X7 -0.008549  0.991487  0.009304 -0.919  0.35816
## X8  0.072327  1.075006  0.232133  0.312  0.75536
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##    exp(coef) exp(-coef) lower .95 upper .95
## X1    1.3363     0.7483    0.8903    2.0058
## X2    0.6706     1.4913    0.3853    1.1669
## X3    1.5791     0.6333    0.9370    2.6611
## X4    2.2005     0.4544    1.2159    3.9824
## X5    0.9679     1.0332    0.9575    0.9784
## X6    0.9999     1.0001    0.9822    1.0180
## X7    0.9915     1.0086    0.9736    1.0097
## X8    1.0750     0.9302    0.6821    1.6943
##
## Concordance= 0.736  (se = 0.03 )
## Rsquare= 0.361   (max possible= 0.999 )
## Likelihood ratio test= 61.41  on 8 df,   p=2.464e-10
## Wald test            = 61.65  on 8 df,   p=2.212e-10
## Score (logrank) test = 65.92  on 8 df,   p=3.178e-11
```

H0: all coefficients=0 or the set of variables is not associated with survival; H1: at least one coefficients is not 0 or one variables in the set is associated with survival.

Model LR chi-square = 61.41 on 8 df, p=2.464e-10. Since the p-value is less than 0.05, with 95% confidence we reject the H0, and conclude that at least one variables in the set is associated with survival in the follow-up study(chi-square = 61.41, p=2.464e-10).

We can also get the below estimated coefficients for each variable, and their p-values.

```
    coef exp(coef)  se(coef)      z Pr(>|z|)
```

X1 0.289936 1.336342 0.207210 1.399 0.16174
X2 -0.399628 0.670570 0.282663 -1.414 0.15742
X3 0.456859 1.579106 0.266273 1.716 0.08621
X4 0.788672 2.200471 0.302668 2.606 0.00917

X5 -0.032622 0.967905 0.005505 -5.926 3.11e-09

X6 -0.000092 0.999908 0.009125 -0.010 0.99196
X7 -0.008549 0.991487 0.009304 -0.919 0.35816
X8 0.072327 1.075006 0.232133 0.312 0.75536

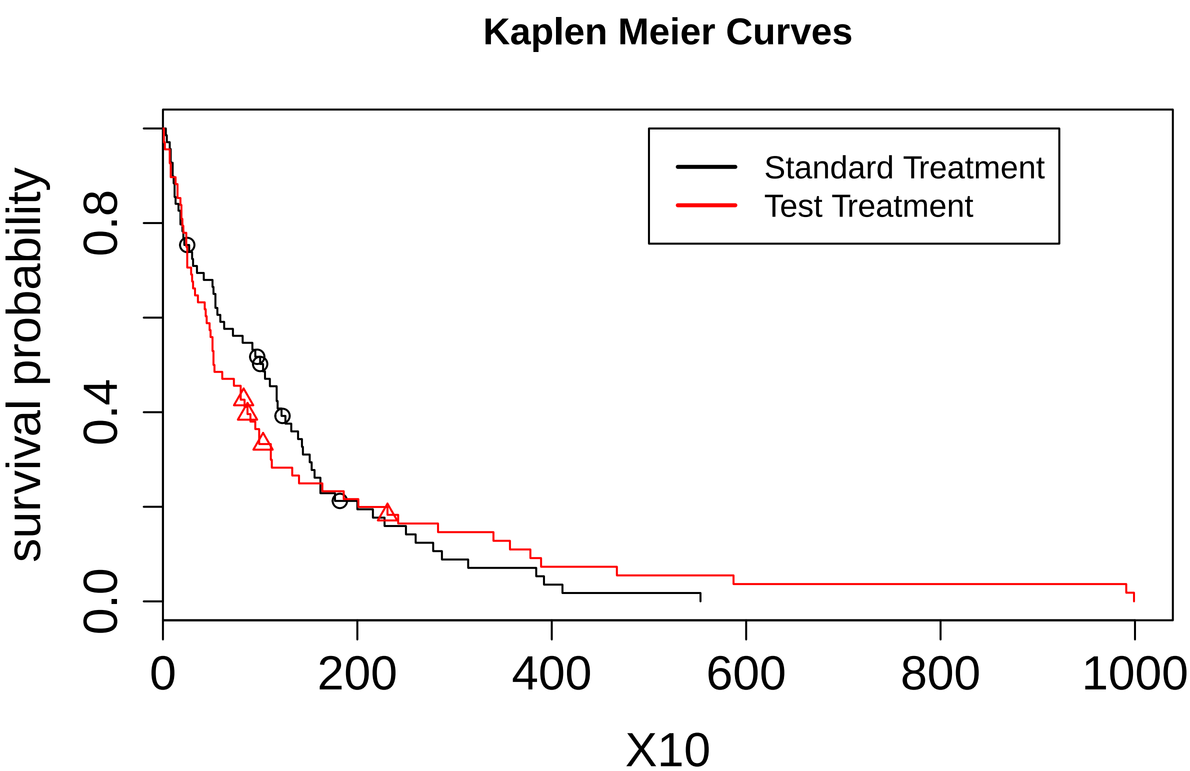
Here we want to test for treatments (X1). H0: coefficient for treatment is 0; H1: coefficient for treatment not equal to 0. LR test: LR chi-square = Full Model chi-square - Model without X1 chi-sauare = 61.41-59.45 = 1.96 with 1 df. p-value = 0.16 > 0.05. With 95% confidence we can not reject the H0, and conclude that coefficient for X1 is 0, so the two treatment groups do not differ with respect to 3-year survival after adjustment for confounders in the follow-up study(chi-square=1.96, p-value = 0.16).

The estimated coefficient for treatment is 0.2899, so HR = exp(coefficient) = exp(0.2899) = 1.34. After adjustment for confounders, the relative risk of dying in the test group compared to the standard group is 1.34.

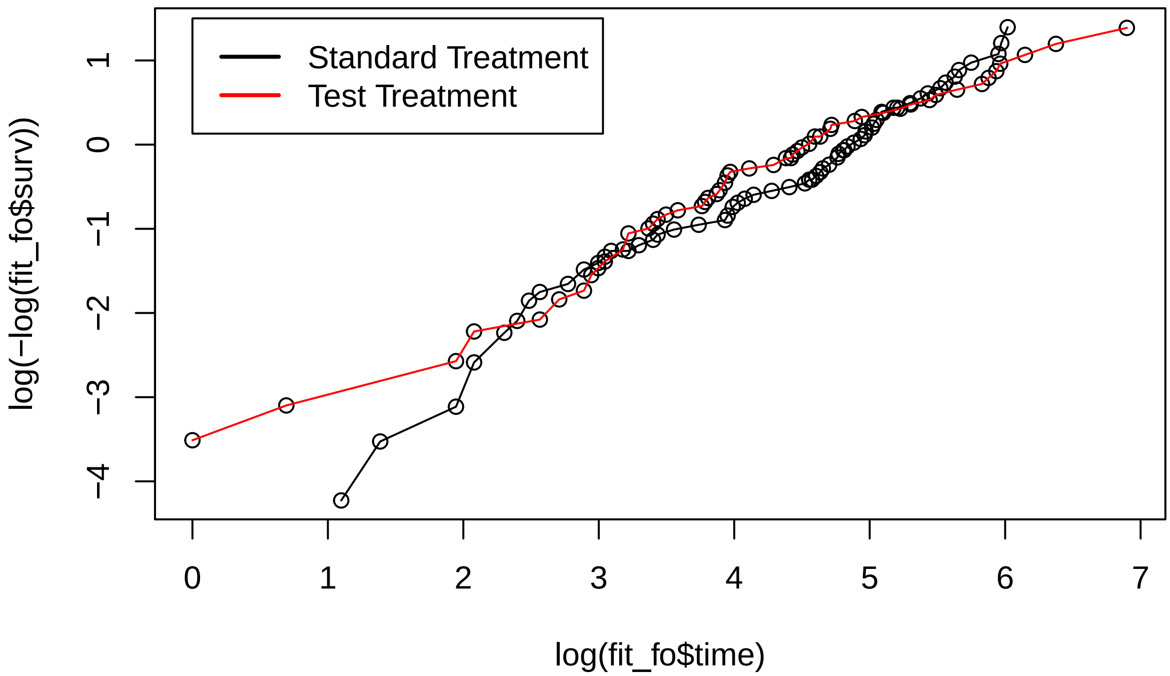
95% confidence interval: (0.89, 2.01). With 95% confidence, we estimate that the true hazard ratio for dying lies between 0.89 and 2.01. This confidence interval includes 1. There is not statistically significant evidence to reject H0 and suggest an increased hazard ratio of dying in the test group after adjustment for confounders.

The proportional hazards assumption are, a. The natural logarithms of h(t|X) plotted against time are parallel for varying values of X, separated by beta. Firsstly, I plot the log(t) and log(-log(S(t)).

```
fit_fo <- survfit(Surv(X10,X11)~X1, data=Lungca)
plot(fit_fo, firstx=0, mark.time=T, mark=c(1,2), col=c(1,2), ylim=c(0,1), xlab="X10", ylab="survival pr
legend(x=500,y=1,legend=c("Standard Treatment","Test Treatment"), col=c(1,2),lwd=2,cex=1)
```




```
plot(log(fit_fo$time),log(-log(fit_fo$surv)))
lines(log(fit_fo$time[1:60]),log(-log(fit_fo$surv[1:60])))
lines(log(fit_fo$time[61:114]),log(-log(fit_fo$surv[61:114])),col=2)
legend(x=0,y=1.5,legend=c("Standard Treatment","Test Treatment"), col=c(1,2),lwd=2,cex=1)
```



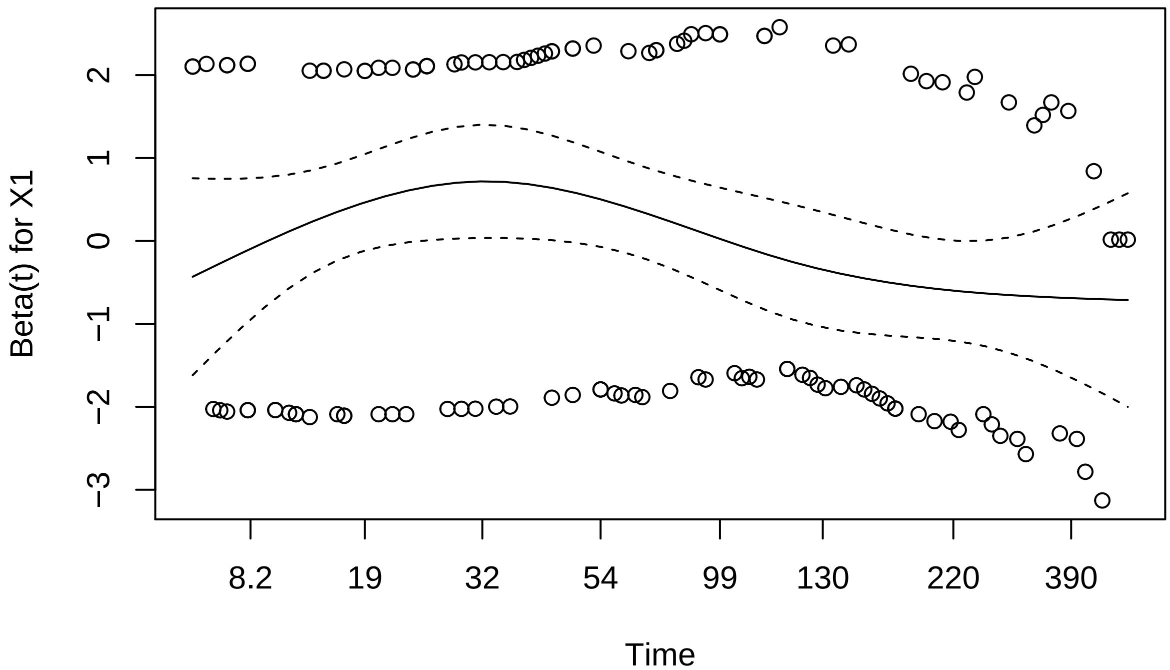
From the plot of log(t) and log(-log(S(t)), we can see that the two line about X1=1 and X1=0 are not parallel, so it does not suggest proportional hazards regression for the assumption of parallelity.

b. The Hazard Ratio does not depend on time. We test the proportionality of hazard with the function, $\beta(t) = \beta + \theta g(t)$ H0: $\theta = 0$ => proportional hazard is OK;
Ha: $\theta \neq 0$ => proportional hazard is not satisfied.

```
test_fo <- cox.zph(fit.crude_fo)
print(test_fo)
```

```
##       rho chisq      p
## X1 -0.159  3.28 0.0701
```

```
plot(test_fo)
```



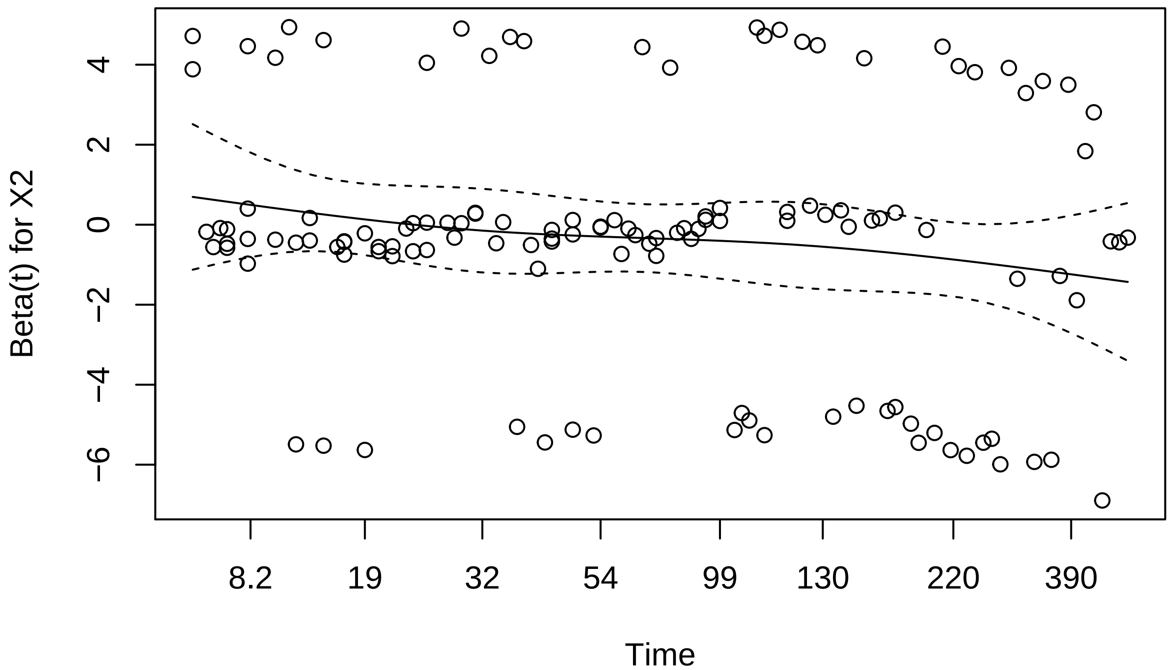
From the output, correlation is -0.159. Conclusions: We fail to reject the null hypothesis since the chi-square is 3.28 with a p-value 0.0701 > 0.05, so this proportional hazards assumption is satisfied for the treatment groups.

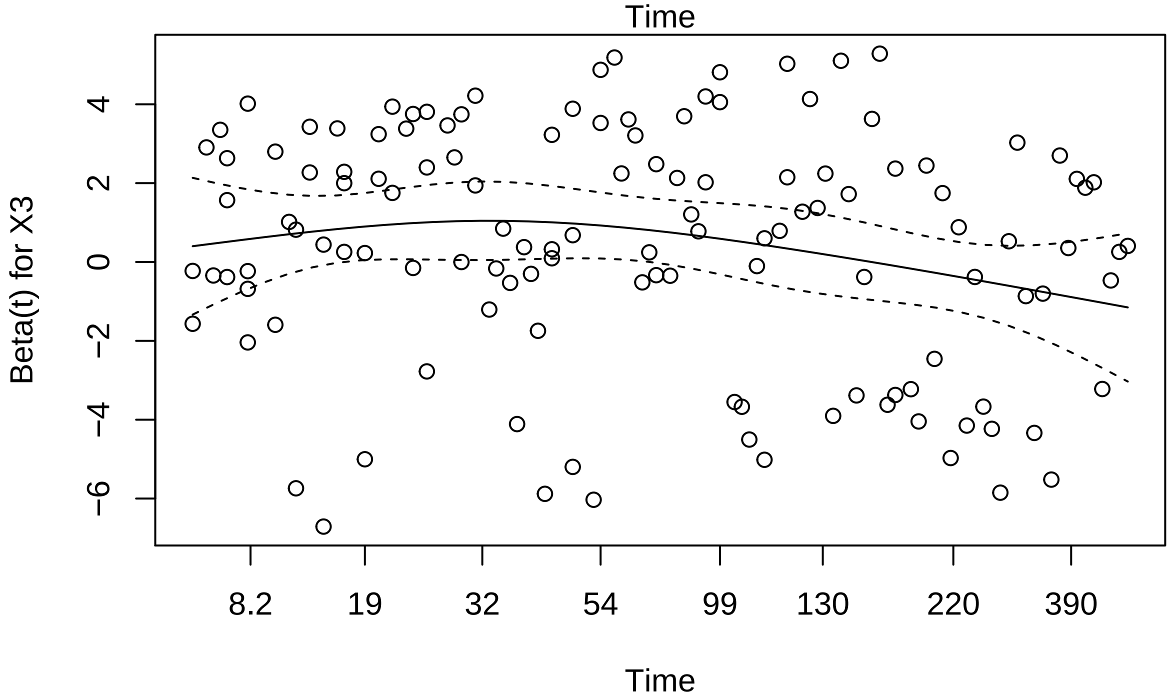
Only confounders:

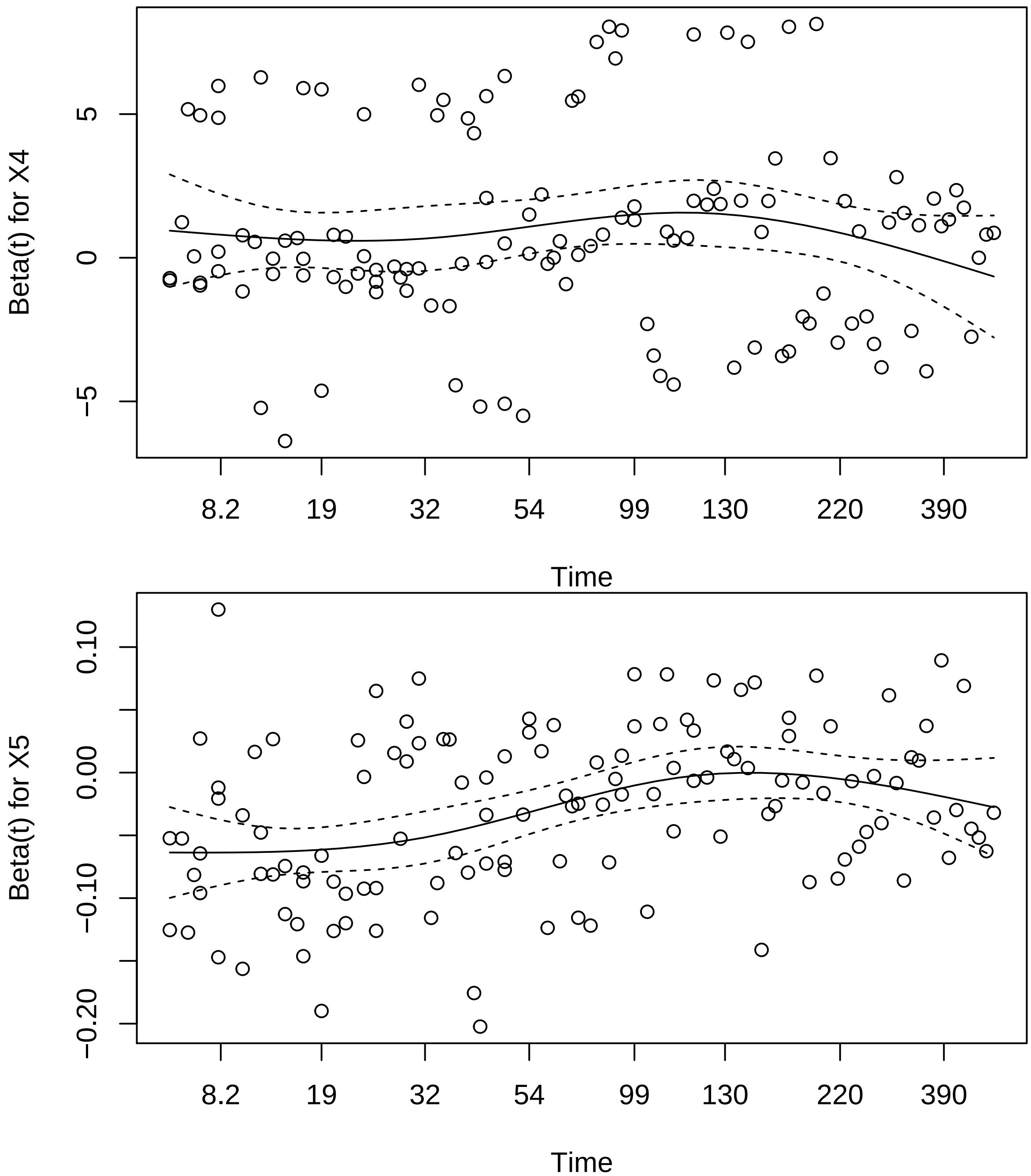
```
test_fo2 <- cox.zph(fit.cov_fo)
print(test_fo2)
```

```
##            rho  chisq        p
## X2     -0.1643  3.437 0.063740
## X3     -0.1688  3.672 0.055339
## X4     -0.0172  0.038 0.845388
## X5      0.3206 14.556 0.000136
## X6      0.1397  2.604 0.106585
## X7      0.2078  6.361 0.011665
## X8     -0.1761  4.332 0.037408
## GLOBAL      NA 28.867 0.000153
```

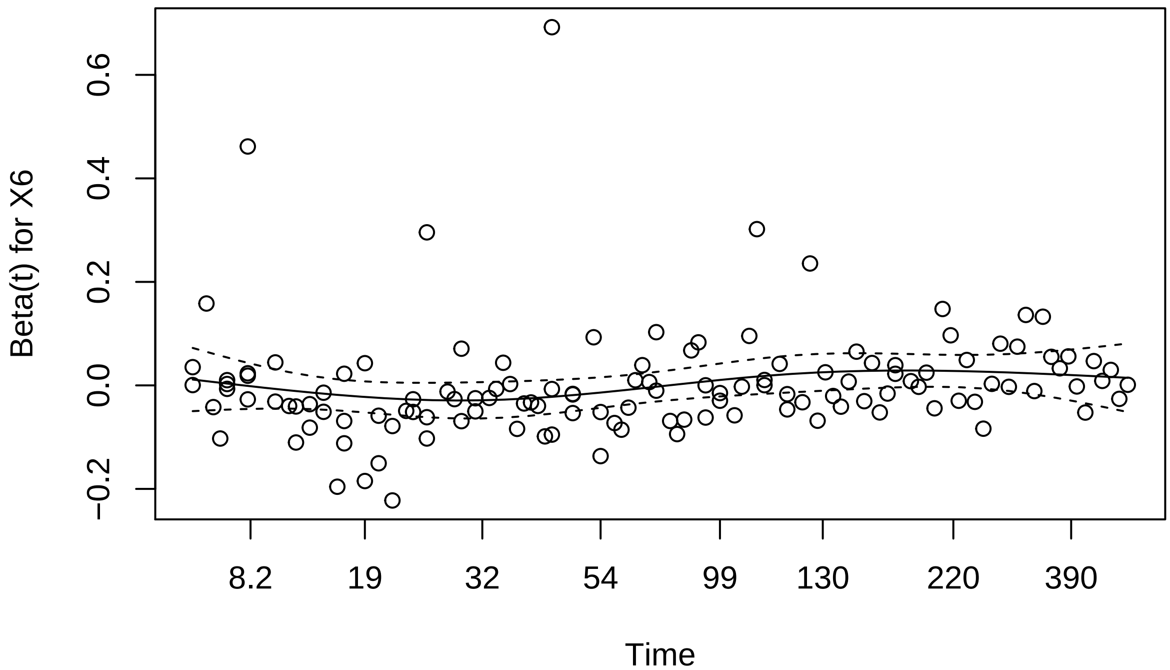
```
plot(test_fo2)
```

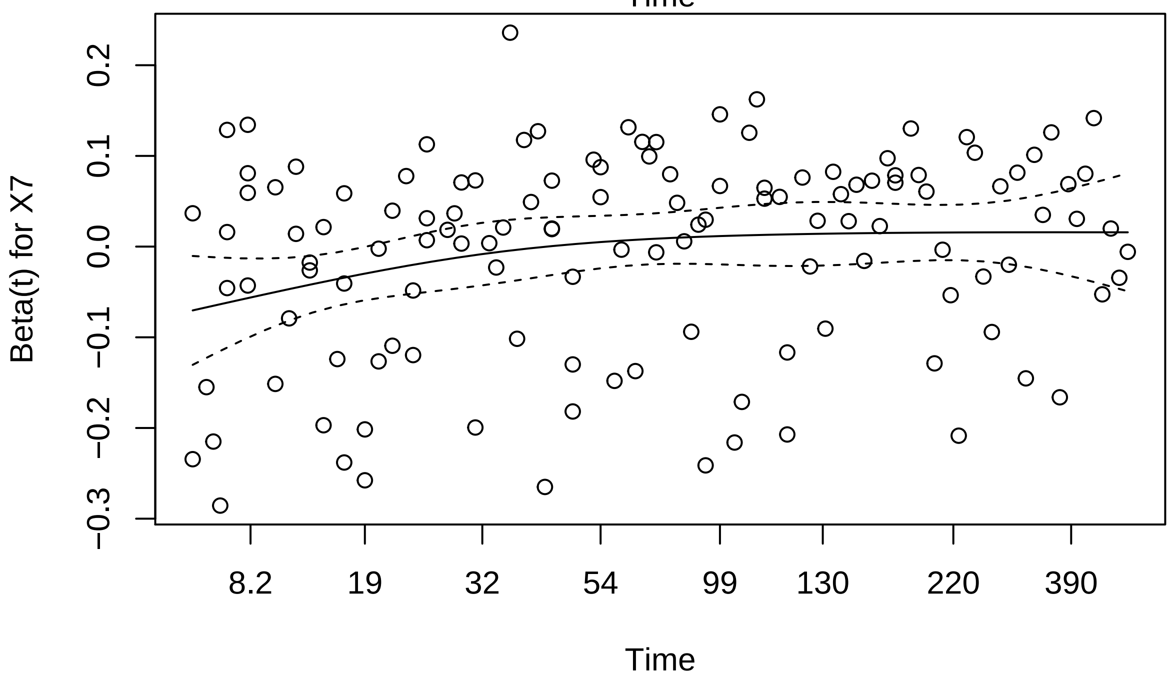

Beta(t) for X2
−6
−4
−2
0
2
4
8.2
19
32
54
99
130
220
390
Time

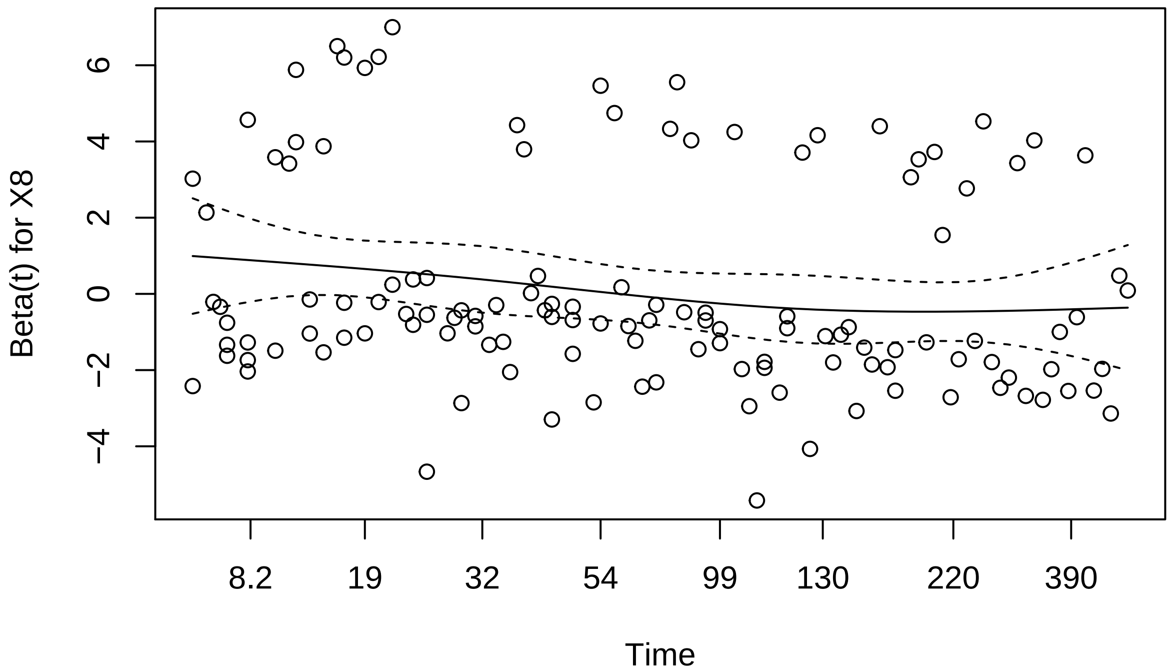

Beta(t) for X3
−6
−4
−2
0
2
4
8.2
19
32
54
99
130
220
390
Time



Beta(t) for X4
5
0
−5
8.2
19
32
54
99
130
220
390
Time
Beta(t) for X5
0.10
0.00
−0.10
−0.20
8.2
19
32
54
99
130
220
390
Time


Beta(t) for X6
−0.2
0.0
0.2
0.4
0.6
8.2
19
32
54
99
130
220
390
Time


Beta(t) for X7
−0.3
−0.2
−0.1
0.0
0.1
0.2
8.2
19
32
54
99
130
220
390
Time



Conclusions: We fail to reject the null hypothesis since the p-value for variables X2, X3, X4 and X6 are larger than 0.05, so the proportional hazards assumption is satisfied for type (X2, X3, X4) and disease duration (X6). However, for the variable performance status (X5), age (X7) and therapy (X8), this proportional hazards assumption is not satisfied since they have small p-values of 0.0001, 0.0117 and 0.0374, respectively .