NLTC_INN

S_16_373

4/1/2022

Load the libraries

```
library(tidyverse)
library(survival)
library(ggfortify)
library(survminer)
library(rms)
library(dynpred)
library(MASS)
library(CPE)
library(scatterplot3d)
library(plotly)
```

Read the data sets

```
Home <- read.csv("~/Research/Research---ST426/Data/NewData/Merge/Home/Home.csv")
Overseas <- read.csv("~/Research/Research---ST426/Data/NewData/Merge/Overseas/Overseas.csv")
```

Merge two data sets

```
data_merge <- rbind(Home , Overseas)
```

Filter the data of NLTC Perera

```
bt12 <- data_merge %>% filter(grepl('NLTC Perera' , Striker))
```

1)Analysis for two different innings

```
NLTC_diff_inn <- survdiff(Surv(cum_balls , wicket) ~ innings, data = bt12)
NLTC_diff_inn
```

```
## Call:
## survdiff(formula = Surv(cum_balls, wicket) ~ innings, data = bt12)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## innings=1 556       32     19.5      7.96      14.5
## innings=2 491       17     29.5      5.27      14.5
##
##  Chisq= 14.5  on 1 degrees of freedom, p= 1e-04
```

The p value is less than 0.05. Therefore there is a significant difference between the innings.

First innings

Filter the first innings

```
NLTC_1 <- bt12 %>% filter(innings == 1)
```

Cox model

```
NLTC_1_cut <- survSplit(NLTC_1 , cut = 23 , end = "cum_balls" , event = "wicket" ,
                        start = "start" , id = "id")

NLTC_1_cut$hv1 <- NLTC_1_cut$cum_runs * (NLTC_1_cut$start < 23)
NLTC_1_cut$hv2 <- NLTC_1_cut$cum_runs * (NLTC_1_cut$start >= 23)

NLTC_1_cut$hv3 <- NLTC_1_cut$cum_singles * (NLTC_1_cut$start < 23)
NLTC_1_cut$hv4 <- NLTC_1_cut$cum_singles * (NLTC_1_cut$start >= 23)


cox_NLTC_1_heav <- coxph(Surv(start,cum_balls , wicket) ~
                           hv1 + hv2 + hv3 + hv4 + ridge(cum_dotBalls) + cluster (id) ,
                         data = NLTC_1_cut)

summary(cox_NLTC_1_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ hv1 + hv2 +
##     hv3 + hv4 + ridge(cum_dotBalls), data = NLTC_1_cut, cluster = id)
##
##   n= 668, number of events= 32
##
##                     coef    se(coef) se2     Chisq  DF p
## hv1                 -0.3041 0.06477  0.06486  22.05 1  2.7e-06
## hv2                 -0.1609 0.04464  0.04962  12.99 1  3.1e-04
## hv3                 -0.6514 0.19825  0.15934  10.80 1  1.0e-03
```

```
## hv4                 -0.8070 0.12395  0.14291  42.39 1  7.5e-11
## ridge(cum_dotBalls) -0.7007 0.06299  0.08718 123.77 1  9.5e-29
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## hv1                    0.7378      1.355    0.6498    0.8376
## hv2                    0.8514      1.175    0.7800    0.9292
## hv3                    0.5213      1.918    0.3535    0.7689
## hv4                    0.4462      2.241    0.3500    0.5689
## ridge(cum_dotBalls)    0.4962      2.015    0.4386    0.5614
##
## Iterations: 2 outer, 10 Newton-Raphson
## Degrees of freedom for terms= 0.9 0.8 0.9 0.8 0.6
## Concordance= 0.974  (se = 0.005 )
## Likelihood ratio test= 135  on 4.1 df,   p=<2e-16
```

Test pH assumptions

```
test_NLTC_1 <- cox.zph(cox_NLTC_1_heav)
test_NLTC_1
```

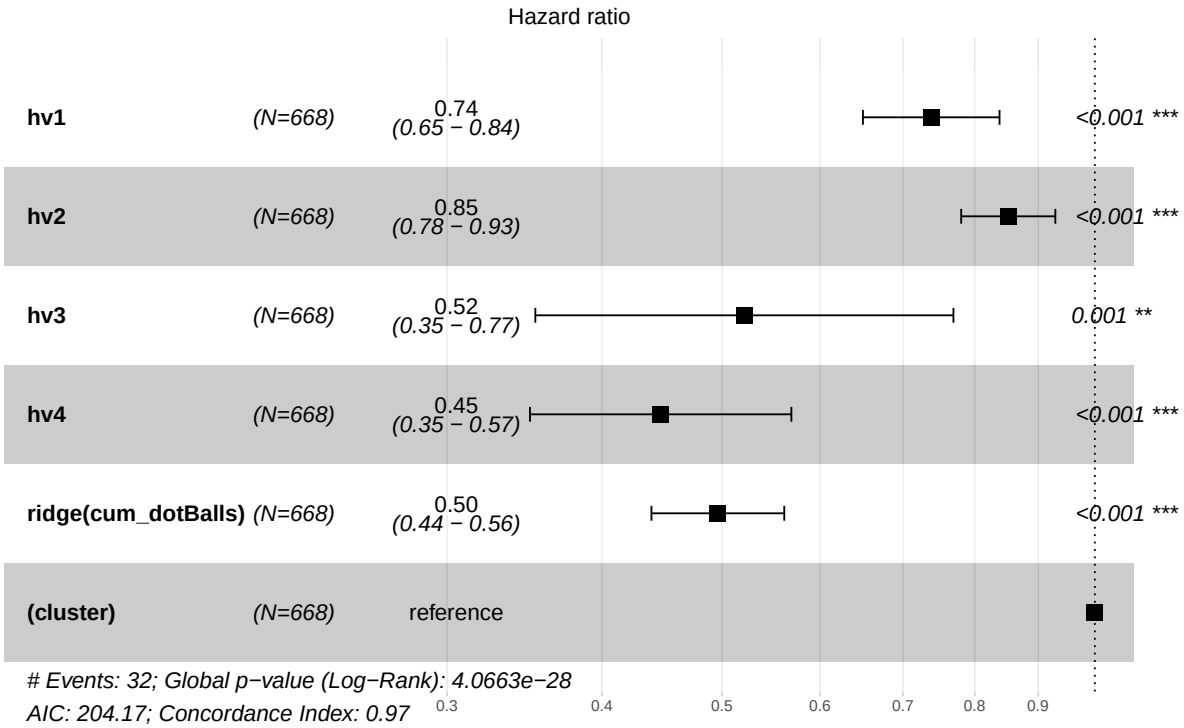
```
##                      chisq   df     p
## hv1                 1.9908 0.91 0.141
## hv2                 0.8646 0.85 0.300
## hv3                 3.0587 0.93 0.072
## hv4                 0.6207 0.84 0.369
## ridge(cum_dotBalls) 0.0232 0.57 0.688
## GLOBAL              6.5947 4.10 0.167
```

Plot the survival curve

```
cox_plot_NLTC_1 <- ggadjustedcurves(cox_NLTC_1_heav , data = NLTC_1_cut , size = 2,
                      ggtheme = theme_survminer()) +
                      geom_hline(yintercept = c(0.25,0.5,0.75) , linetype = "dashed") +
                      ggtitle("Survival curve for NLTC Perera for the first innings") +
                      theme(axis.title = element_text(size = 15),
                            axis.text = element_text(size = 14))
```

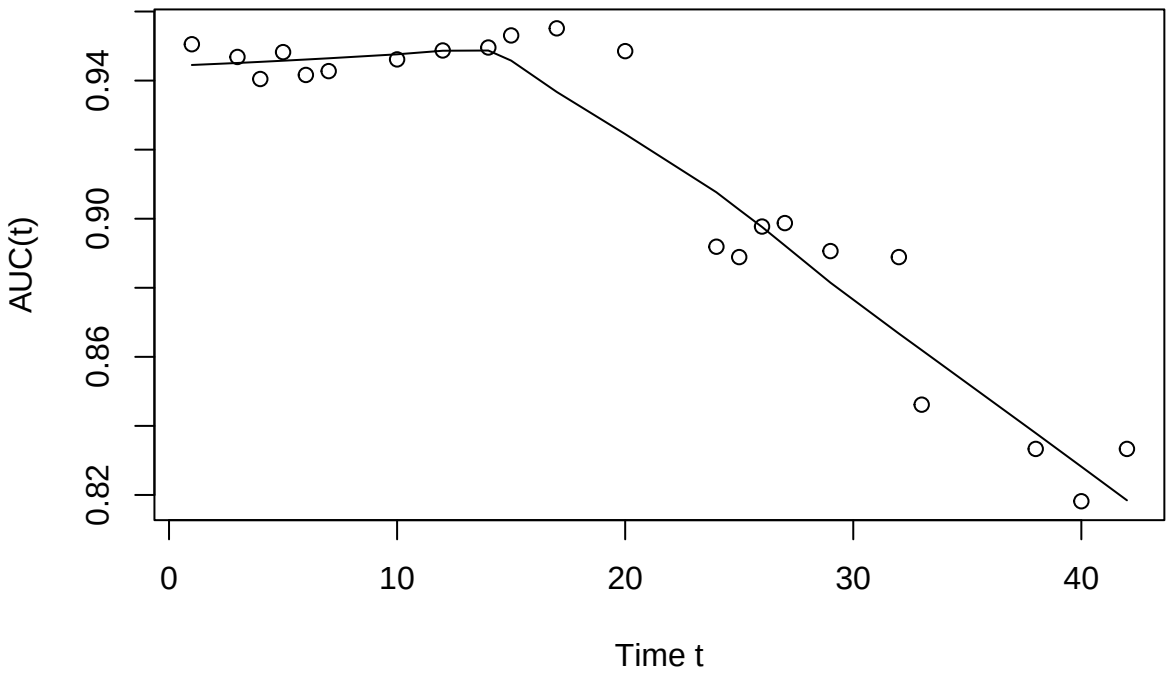
coeffient plot

```
coef_plot_NLTC_1 <- ggforest(cox_NLTC_1_heav , data = NLTC_1_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_NLTC_1
```



Area under the curve

```
AUC(Surv(start,cum_balls , wicket) ~
      hv1 + hv2 + hv3 + hv4 + ridge(cum_dotBalls) + cluster (id) , data = NLTC_1_cut)
```




```
## $AUCt
##    time       AUC
## 1     1 0.9505247
## 2     3 0.9468439
## 3     4 0.9404553
## 4     5 0.9482440
## 5     6 0.9416342
## 6     7 0.9427403
## 7    10 0.9461358
## 8    12 0.9487179
## 9    14 0.9495798
## 10   15 0.9530792
## 11   17 0.9551282
## 12   20 0.9485294
## 13   24 0.8918919
## 14   25 0.8888889
## 15   26 0.8977273
## 16   27 0.8987342
## 17   29 0.8906250
## 18   32 0.8888889
## 19   33 0.8461538
## 20   38 0.8333333
## 21   40 0.8181818
## 22   42 0.8333333
##
## $AUC
```

```
## [1] 0.9414198
```

Second Innings

Filter the second innings

```
NLTC_2 <- bt12 %>% filter(innings == 2)
```

Cox model

```
NLTC_2_cut <- survSplit(NLTC_2 , cut = 23 , end = "cum_balls" , event = "wicket" ,
                        start = "start" , id = "id")

NLTC_2_cut$hv1 <- NLTC_2_cut$cum_runs * (NLTC_2_cut$start < 23)
NLTC_2_cut$hv2 <- NLTC_2_cut$cum_runs * (NLTC_2_cut$start >= 23)

NLTC_2_cut$hv3 <- NLTC_2_cut$cum_singles * (NLTC_2_cut$start < 23)
NLTC_2_cut$hv4 <- NLTC_2_cut$cum_singles * (NLTC_2_cut$start >= 23)


cox_NLTC_2_heav <- coxph(Surv(start,cum_balls , wicket) ~
                           hv1 + hv2+ hv3 + hv4 +ridge(cum_dotBalls) + cluster (id) ,
                         data = NLTC_2_cut)
summary(cox_NLTC_2_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ hv1 + hv2 +
##     hv3 + hv4 + ridge(cum_dotBalls), data = NLTC_2_cut, cluster = id)
##
##   n= 670, number of events= 17
##
##                       coef     se(coef) se2     Chisq DF p
## hv1                   -0.08420 0.03493  0.04875  5.81 1  1.6e-02
## hv2                   -0.12226 0.04926  0.04422  6.16 1  1.3e-02
## hv3                   -0.65674 0.18970  0.17317 11.98 1  5.4e-04
## hv4                   -0.07621 0.26309  0.20211  0.08 1  7.7e-01
## ridge(cum_dotBalls) -0.28765 0.05456  0.05222 27.79 1  1.3e-07
##
##                       exp(coef) exp(-coef) lower .95 upper .95
## hv1                      0.9192      1.088    0.8584    0.9844
## hv2                      0.8849      1.130    0.8035    0.9746
## hv3                      0.5185      1.928    0.3575    0.7521
## hv4                      0.9266      1.079    0.5533    1.5518
## ridge(cum_dotBalls)      0.7500      1.333    0.6740    0.8347
##
## Iterations: 2 outer, 10 Newton-Raphson
## Degrees of freedom for terms= 1.0 0.9 1.0 1.0 0.4
## Concordance= 0.967  (se = 0.01 )
## Likelihood ratio test= 71.91  on 4.32 df,   p=1e-14
```

Test pH assumptions

```
test_NLTC_2 <- cox.zph(cox_NLTC_2_heav)
test_NLTC_2
```

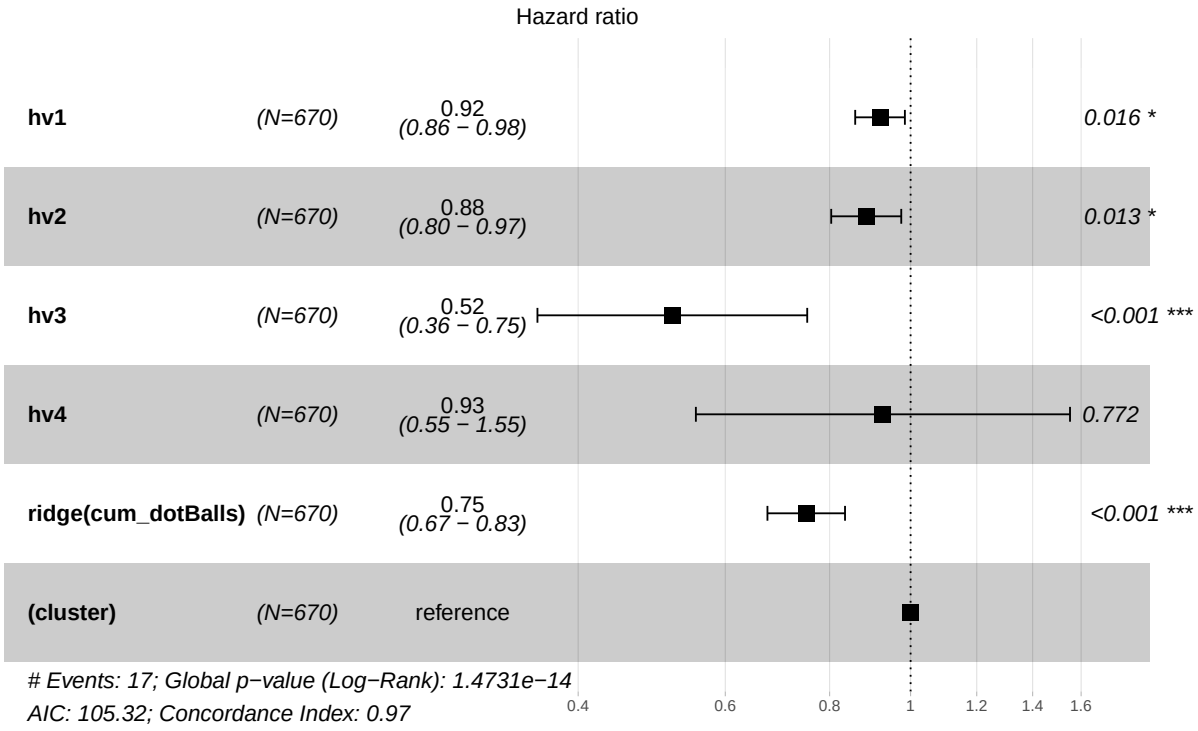
```
##                         chisq   df     p
## hv1                  1.07e-04 0.99 0.991
## hv2                  8.97e-01 0.95 0.326
## hv3                  2.43e-02 0.99 0.874
## hv4                  3.56e+00 0.98 0.057
## ridge(cum_dotBalls) 2.47e-05 0.41 0.891
## GLOBAL               8.64e+00 4.32 0.086
```

Plot the survival curve

```
cox_plot_NLTC_2 <- ggadjustedcurves(cox_NLTC_2_heav , data = NLTC_2_cut , size = 2,
                         ggtheme = theme_survminer()) +
                         geom_hline(yintercept = c(0.25,0.5,0.75) , linetype = "dashed") +
                         ggtitle("Survival curve for NLTC Perera for the second innings") +
                         theme(axis.title = element_text(size = 15),
                               axis.text = element_text(size = 14))
```

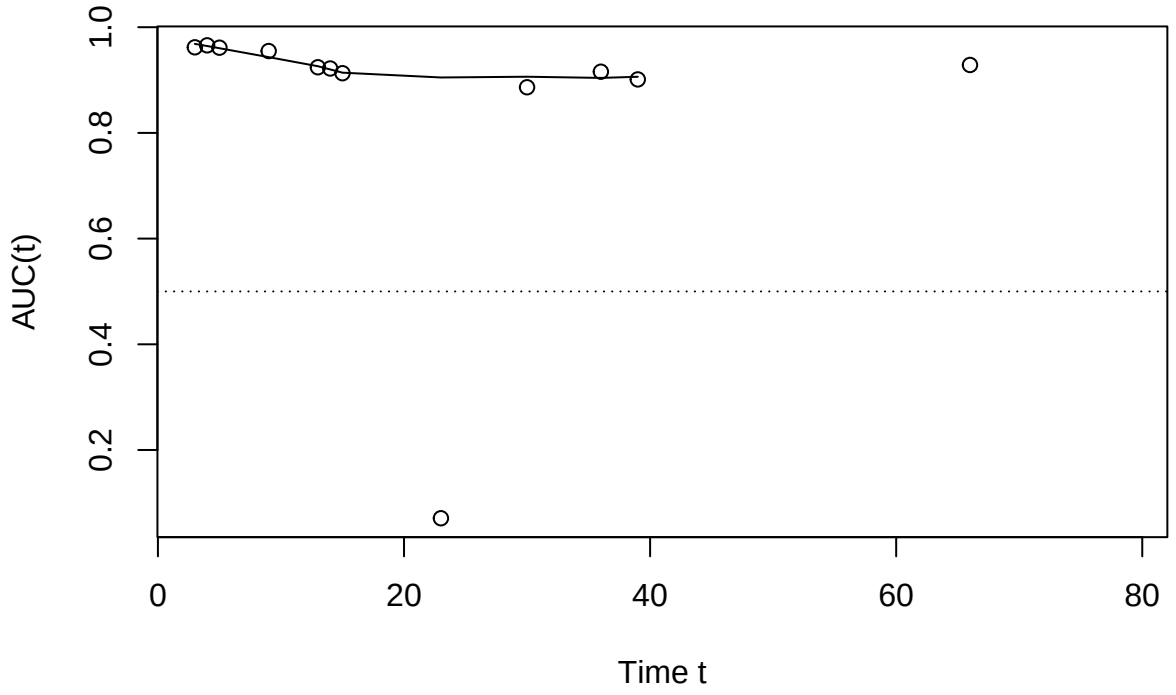
coeffient plot

```
coef_plot_NLTC_2 <- ggforest(cox_NLTC_2_heav , data = NLTC_2_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_NLTC_2
```



Area under the curve

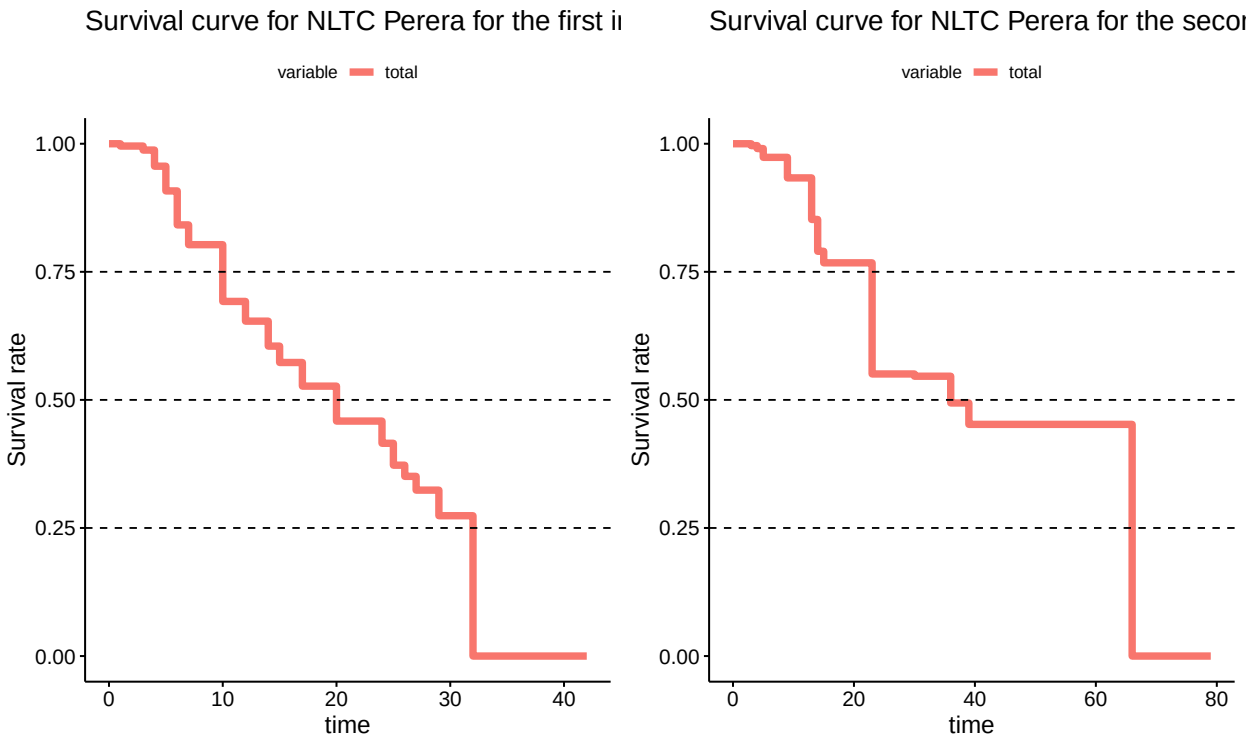
```
AUC(Surv(start,cum_balls , wicket) ~
      hv1 + hv2+ hv3 + hv4 +ridge(cum_dotBalls) + cluster (id) , data = NLTC_2_cut)
```




```
## $AUCt
##    time        AUC
## 1     3 0.96196513
## 2     4 0.96568627
## 3     5 0.96127946
## 4     9 0.95488722
## 5    13 0.92421053
## 6    14 0.92190889
## 7    15 0.91294643
## 8    23 0.07084469
## 9    30 0.88636364
## 10   36 0.91578947
## 11   39 0.90123457
## 12   66 0.92857143
## 13   79        NaN
##
## $AUC
## [1] 0.8703287
```

Survival curve

```
NLTCSurvPlotInn <- ggarrange(cox_plot_NLTC_1,cox_plot_NLTC_2,ncol = 2)
NLTCSurvPlotInn
```



Upload to texStudio

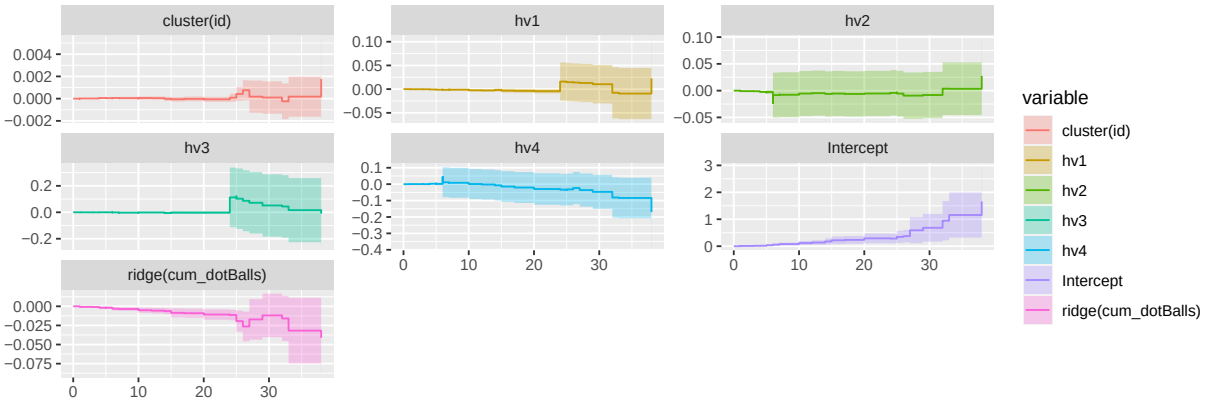
```
filepath <- "~/Research/Thesis_S16373/images/"
postscript(file = paste0(filepath,"NLTCSurvPlotInn.eps"),width = 10,height = 7,horizontal = FALSE)
NLTCSurvPlotInn
dev.off()
```

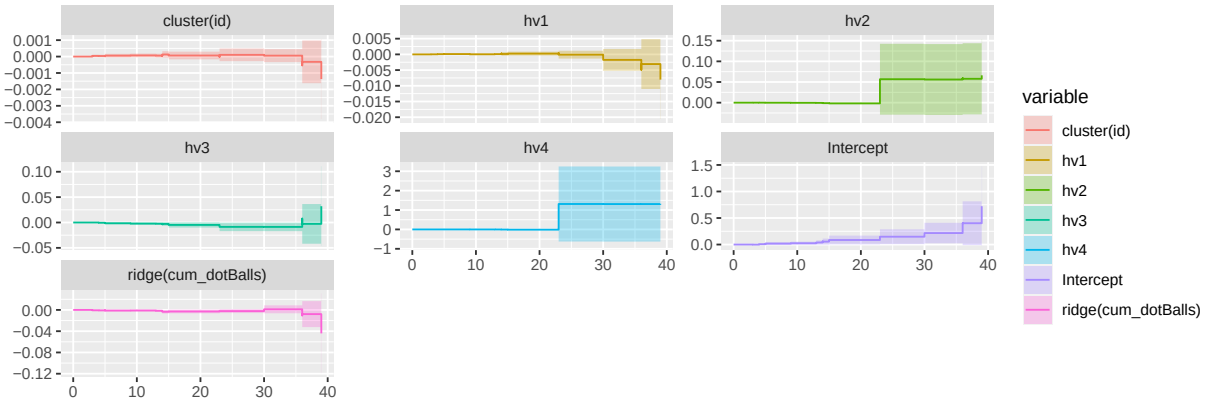
Aalen model

```
NLTC_p1 <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
        hv1 + hv2 +hv3 +hv4 + ridge(cum_dotBalls) + cluster (id) , data = NLTC_1_cut))

NLTC_p2 <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
        hv1 + hv2 +hv3 +hv4 + ridge(cum_dotBalls) + cluster (id) , data = NLTC_2_cut))

NLTCAalenInn <- ggarrange(NLTC_p1,NLTC_p2, nrow = 2)
NLTCAalenInn
```






```
filepath <- "~/Research/Thesis_S16373/images/"
postscript(file = paste0(filepath,"NLTCAalenInn.eps"),width = 10,height = 7,horizontal = FALSE)
NLTCAalenInn
dev.off()
```