DMS_POS

S_16_373

5/8/2022

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 DM de Silva

```
bt <- data_merge %>% filter(grepl('DM de Silva' , Striker))
```

1)Analysis for different batting position

```r
bt_ps <- bt %>% filter(bat_position == c(5,6,7))

DMS_diff_pos <- survdiff(Surv(cum_balls , wicket) ~ bat_position, data = bt_ps)
DMS_diff_pos
```

```
## Call:
## survdiff(formula = Surv(cum_balls, wicket) ~ bat_position, data = bt_ps)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## bat_position=5  85        3     2.65    0.0469    0.0725
## bat_position=6 125        2     2.65    0.1589    0.2610
## bat_position=7  79        3     2.70    0.0325    0.0499
##
##  Chisq= 0.3  on 2 degrees of freedom, p= 0.9
```

The p value is greater than 0.05. Therefore we cannot see a significant difference between the batting positions 5,6 and 7.

batting position 5

Filter the first bat_position

```r
DMS_P_5<- bt %>% filter(bat_position == 5)
```

Cox model

```r
DMS_P_5_cut <- survSplit(DMS_P_5 , cut = 30 , end = "cum_balls" , event = "wicket" , start = "start" ,

DMS_P_5_cut$hv1 <- DMS_P_5_cut$cum_runs  * (DMS_P_5_cut$start < 30)
DMS_P_5_cut$hv2 <- DMS_P_5_cut$cum_runs  * (DMS_P_5_cut$start >= 30)

cox_DMS_P_5_heav <- coxph(Surv(start,cum_balls , wicket) ~
                            cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id
                          data = DMS_P_5_cut)

summary(cox_DMS_P_5_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ cum_singles +
##     hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls), data = DMS_P_5_cut,
##     cluster = id)
##
##   n= 308, number of events= 10
##
##                       coef    se(coef) se2    Chisq DF p
## cum_singles         -0.8116 0.2137   0.2816 14.42 1  1.5e-04
## hv1                 -0.6365 0.1884   0.2886 11.41 1  7.3e-04
## hv2                  0.0604 0.1693   0.3945  0.13 1  7.2e-01
## cum_boundaries       0.4270 1.5114   1.5041  0.08 1  7.8e-01
## ridge(cum_dotBalls) -0.8309 0.1759   0.2300 22.31 1  2.3e-06
```

```
##
##                       exp(coef) exp(-coef) lower .95 upper .95
## cum_singles              0.4441     2.2515   0.29214    0.6752
## hv1                      0.5292     1.8898   0.36575    0.7656
## hv2                      1.0623     0.9414   0.76232    1.4802
## cum_boundaries           1.5327     0.6525   0.07924   29.6438
## ridge(cum_dotBalls)      0.4357     2.2953   0.30863    0.6150
##
## Iterations: 5 outer, 22 Newton-Raphson
## Degrees of freedom for terms= 0.9 0.9 1.0 1.0 0.5
## Concordance= 0.98  (se = 0.009 )
## Likelihood ratio test= 56.06  on 4.3 df,   p=3e-11
```

Test pH assumptions

```
test_DMS_P_5 <- cox.zph(cox_DMS_P_5_heav)
test_DMS_P_5
```

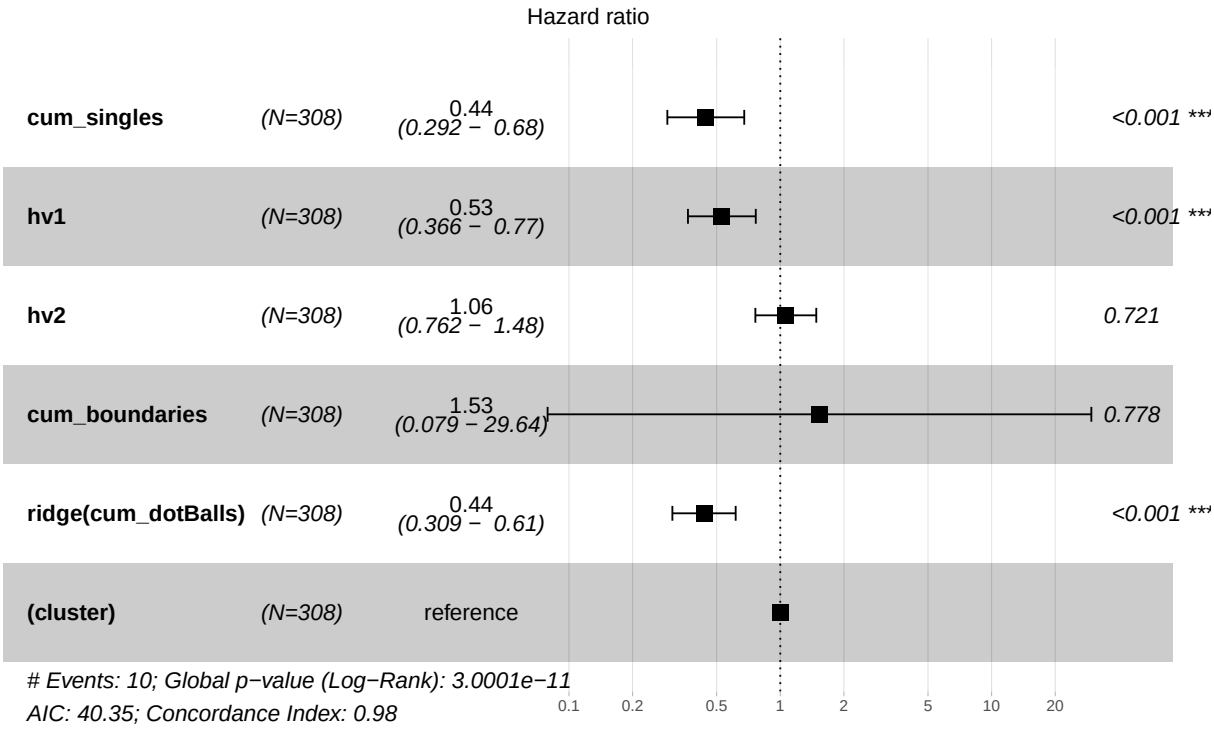
```
##                         chisq   df    p
## cum_singles           1.37265 0.92 0.22
## hv1                   0.61356 0.91 0.40
## hv2                   0.36796 1.00 0.54
## cum_boundaries        0.15329 0.95 0.68
## ridge(cum_dotBalls) 0.00196 0.51 0.81
## GLOBAL                1.56081 4.30 0.85
```

Plot the survival curve

```
cox_plot_DMS_P_5 <- ggadjustedcurves(cox_DMS_P_5_heav , data = DMS_P_5_cut , size = 2,
                        ggtheme =theme_survminer()) + geom_hline(yintercept = c(0.25,0.5,0.75) ,
                        linetype = "dashed") +ggtitle("Survival curve for DM de Silva for the 5th
                        batting position") + theme(axis.title = element_text(size = 15),axis.text =
                        element_text(size = 14))
```

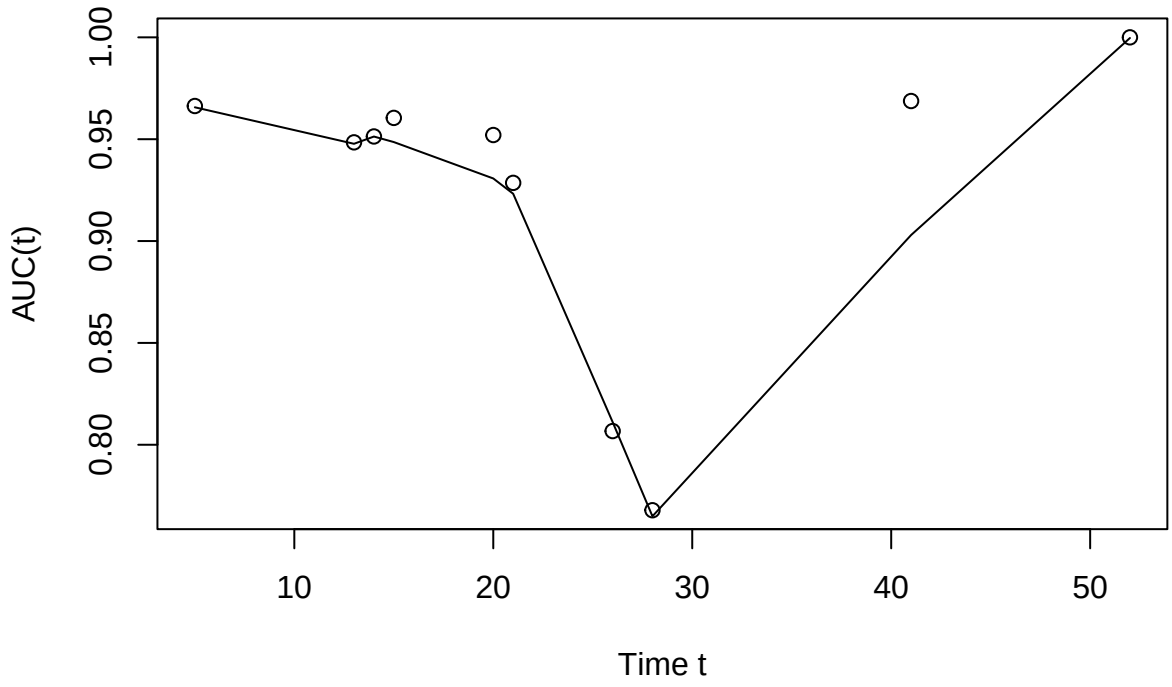
coeffient plot

```
coef_plot_DMS_P_5 <- ggforest(cox_DMS_P_5_heav , data = DMS_P_5_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_DMS_P_5
```



Area under the curve

```
AUC(Surv(start,cum_balls , wicket) ~
      cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = DMS_P_5_cut
```




```
## $AUCt
##    time       AUC
## 1     5 0.9662921
## 2    13 0.9484536
## 3    14 0.9513514
## 4    15 0.9604520
## 5    20 0.9520548
## 6    21 0.9285714
## 7    26 0.8067227
## 8    28 0.7678571
## 9    41 0.9687500
## 10   52 1.0000000
##
## $AUC
## [1] 0.9267241
```

batting position 6

Filter the first bat_position

```
DMS_P_6<- bt %>% filter(bat_position == 6)
```

Cox model

```
DMS_P_6_cut <- survSplit(DMS_P_6 , cut = 30 , end = "cum_balls" , event = "wicket" , start = "start" ,

DMS_P_6_cut$hv1 <- DMS_P_6_cut$cum_runs  * (DMS_P_6_cut$start < 30)
DMS_P_6_cut$hv2 <- DMS_P_6_cut$cum_runs  * (DMS_P_6_cut$start >= 30)

cox_DMS_P_6_heav <- coxph(Surv(start,cum_balls , wicket) ~
                            cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id
                          data = DMS_P_6_cut)

summary(cox_DMS_P_6_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ cum_singles +
##     hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls), data = DMS_P_6_cut,
##     cluster = id)
##
##   n= 481, number of events= 13
##
##                       coef    se(coef) se2    Chisq DF p
## cum_singles         -0.3886 0.31262  0.4178  1.55 1  2.1e-01
## hv1                 -1.0882 0.26448  0.3913 16.93 1  3.9e-05
## hv2                 -0.6775 0.12980  0.2435 27.24 1  1.8e-07
## cum_boundaries       4.3455 0.90895  1.5090 22.86 1  1.7e-06
## ridge(cum_dotBalls) -0.4052 0.09713  0.1168 17.40 1  3.0e-05
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## cum_singles            0.6780    1.47498    0.3674    1.2512
## hv1                    0.3368    2.96905    0.2006    0.5656
## hv2                    0.5079    1.96889    0.3938    0.6550
## cum_boundaries        77.1285    0.01297   12.9872  458.0521
## ridge(cum_dotBalls)    0.6669    1.49958    0.5513    0.8067
##
## Iterations: 4 outer, 17 Newton-Raphson
## Degrees of freedom for terms= 1.0 1.0 1.0 1.0 0.4
## Concordance= 0.982  (se = 0.006 )
## Likelihood ratio test= 69.23  on 4.38 df,   p=6e-14
```

Test pH assumptions

```
test_DMS_P_6 <- cox.zph(cox_DMS_P_6_heav)
test_DMS_P_6
```

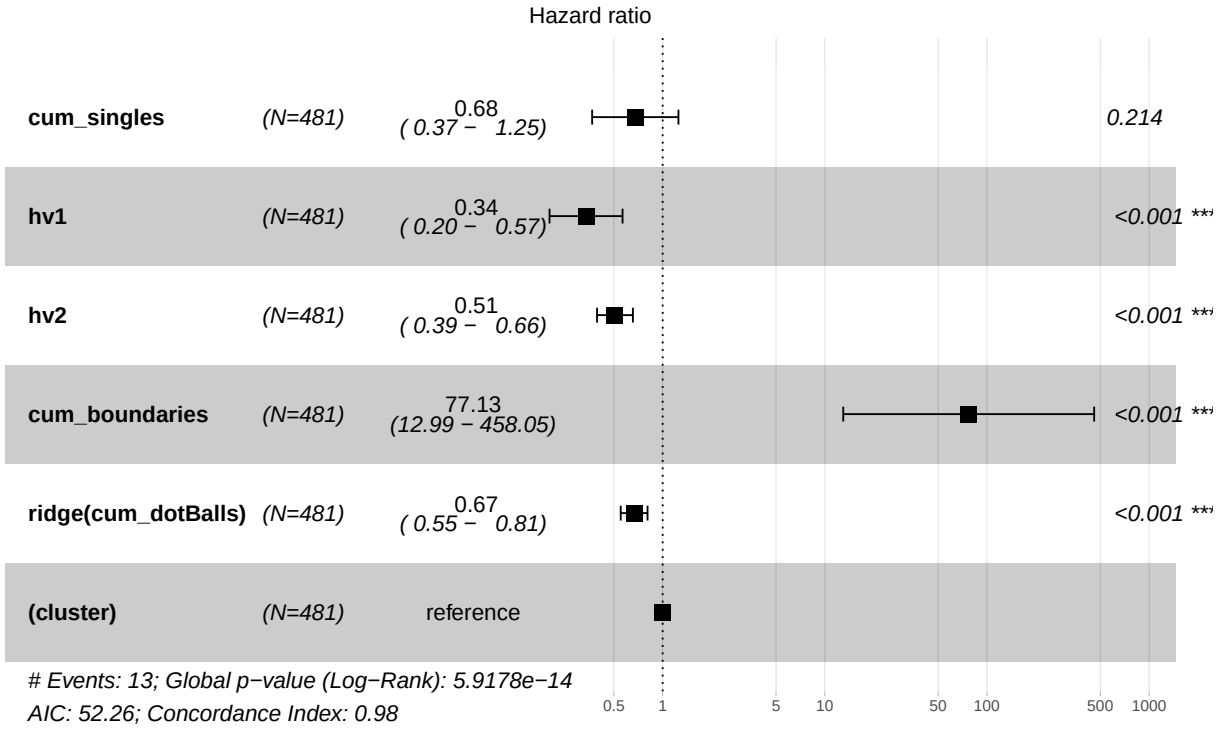
```
##                       chisq   df    p
## cum_singles          1.82484 0.99 0.17
## hv1                  0.36732 1.00 0.54
## hv2                  0.61602 0.97 0.42
## cum_boundaries       1.30194 1.00 0.25
## ridge(cum_dotBalls)  0.00299 0.43 0.73
## GLOBAL              -8.14787 4.38 1.00
```

Plot the survival curve

```
cox_plot_DMS_P_6 <- ggadjustedcurves(cox_DMS_P_6_heav , data = DMS_P_6_cut , size = 2,
                        ggtheme =theme_survminer()) + geom_hline(yintercept = c(0.25,0.5,0.75) ,
                        linetype = "dashed") +ggtitle("Survival curve for DM de Silva for the 6th
                        batting position") + theme(axis.title = element_text(size = 15),axis.text =
                        element_text(size = 14))
```

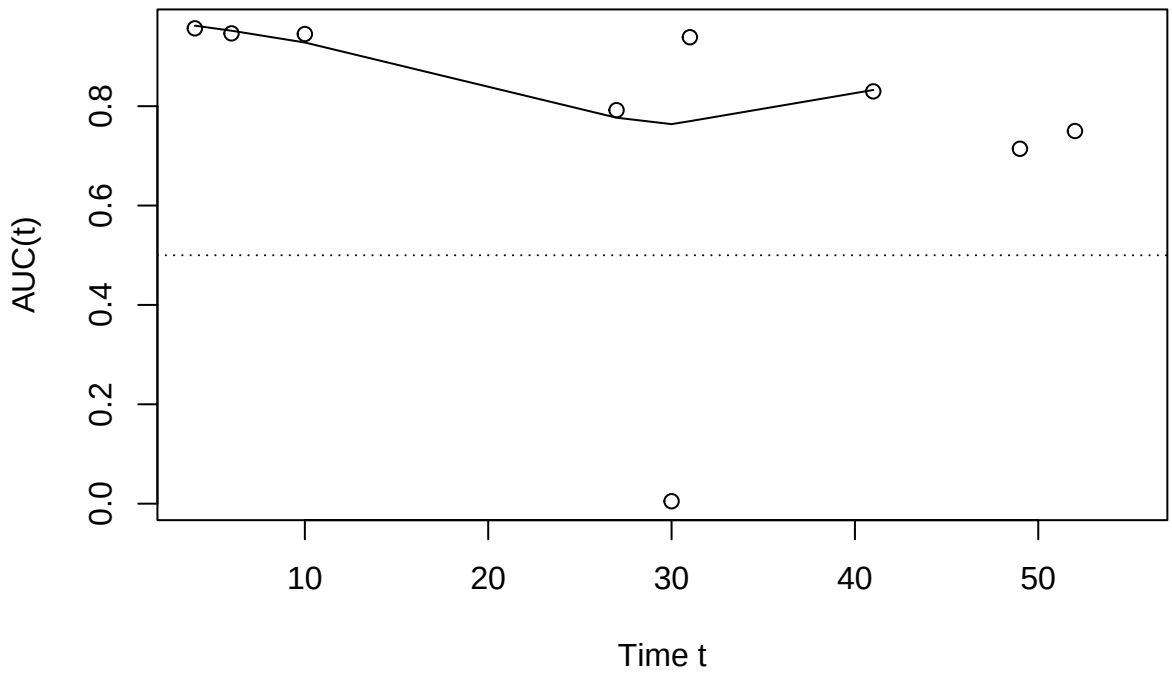
coeffient plot

```
coef_plot_DMS_P_6 <- ggforest(cox_DMS_P_6_heav , data = DMS_P_6_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_DMS_P_6
```



Area under the curve

```
AUC(Surv(start,cum_balls , wicket) ~
      cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = DMS_P_6_cut
```




```
## $AUCt
##    time         AUC
## 1     4 0.956621005
## 2     6 0.946472019
## 3    10 0.945205479
## 4    27 0.792035398
## 5    30 0.004901961
## 6    31 0.938775510
## 7    41 0.829787234
## 8    49 0.714285714
## 9    52 0.750000000
## 10   55         NaN
##
## $AUC
## [1] 0.8173769
```

7th batting position

Filter the second bat_position

```r
DMS_P_7 <- bt %>% filter(bat_position == 7)
```

Cox model

```r
DMS_P_7_cut <- survSplit(DMS_P_7 , cut = 30 , end = "cum_balls" , event = "wicket" , start = "start" ,

DMS_P_7_cut$hv1 <- DMS_P_7_cut$cum_runs  * (DMS_P_7_cut$start < 30)
DMS_P_7_cut$hv2 <- DMS_P_7_cut$cum_runs  * (DMS_P_7_cut$start >= 30)

cox_DMS_P_7_heav <- coxph(Surv(start,cum_balls , wicket) ~
                            cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id
                          data = DMS_P_7_cut)
summary(cox_DMS_P_7_heav)
```

```
## Call:
## coxph(formula = Surv(start, cum_balls, wicket) ~ cum_singles +
##     hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls), data = DMS_P_7_cut,
##     cluster = id)
##
##   n= 322, number of events= 5
##
##                       coef     se(coef) se2     Chisq DF p
## cum_singles         -0.67167 0.45615   0.7074  2.17 1  0.14000
## hv1                  0.01691 0.07616   0.4703  0.05 1  0.82000
## hv2                 -1.56929 0.45005  84.8417 12.16 1  0.00049
## cum_boundaries       1.17774 1.01477   2.4873  1.35 1  0.25000
## ridge(cum_dotBalls) -0.12894 0.16167   0.1005  0.64 1  0.43000
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## cum_singles            0.5109     1.9575   0.20894     1.249
## hv1                    1.0171     0.9832   0.87603     1.181
## hv2                    0.2082     4.8032   0.08618     0.503
## cum_boundaries         3.2470     0.3080   0.44434    23.728
## ridge(cum_dotBalls)    0.8790     1.1376   0.64031     1.207
##
## Iterations: 4 outer, 80 Newton-Raphson
## Degrees of freedom for terms= 1.0 1.0 1.0 1.0 0.5
## Concordance= 0.914  (se = 0.067 )
## Likelihood ratio test= 21.81  on 4.53 df,   p=4e-04
```

Test pH assumptions

```r
test_DMS_P_7 <- cox.zph(cox_DMS_P_7_heav)
test_DMS_P_7
```

```
##                         chisq   df    p
## cum_singles          1.01e-01 0.98 0.74
## hv1                  9.43e-01 1.00 0.33
```

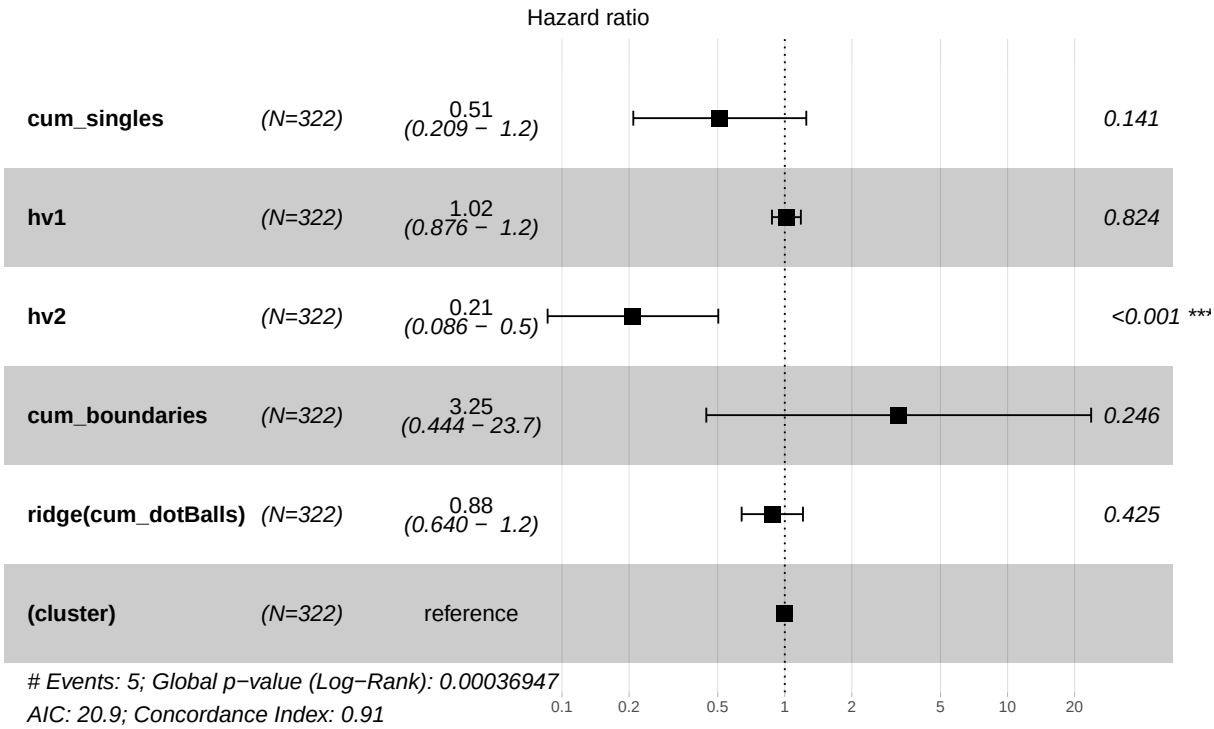
```
## hv2                   1.88e-01 1.00 0.66
## cum_boundaries       -3.76e-01 1.00 1.00
## ridge(cum_dotBalls)   4.66e-03 0.55 0.79
## GLOBAL               -4.59e+05 4.53 1.00
```

Plot the survival curve

```
cox_plot_DMS_P_7 <- ggadjustedcurves(cox_DMS_P_7_heav , data = DMS_P_7_cut , size = 2,ggtheme =
                        theme_survminer()) + geom_hline(yintercept = c(0.25,0.5,0.75) ,
                        linetype = "dashed") +ggtitle("Survival curve for DM de Silva for
                        the 7th batting position") + theme(axis.title = element_text(size = 15),
                        axis.text = element_text(size = 14))
```

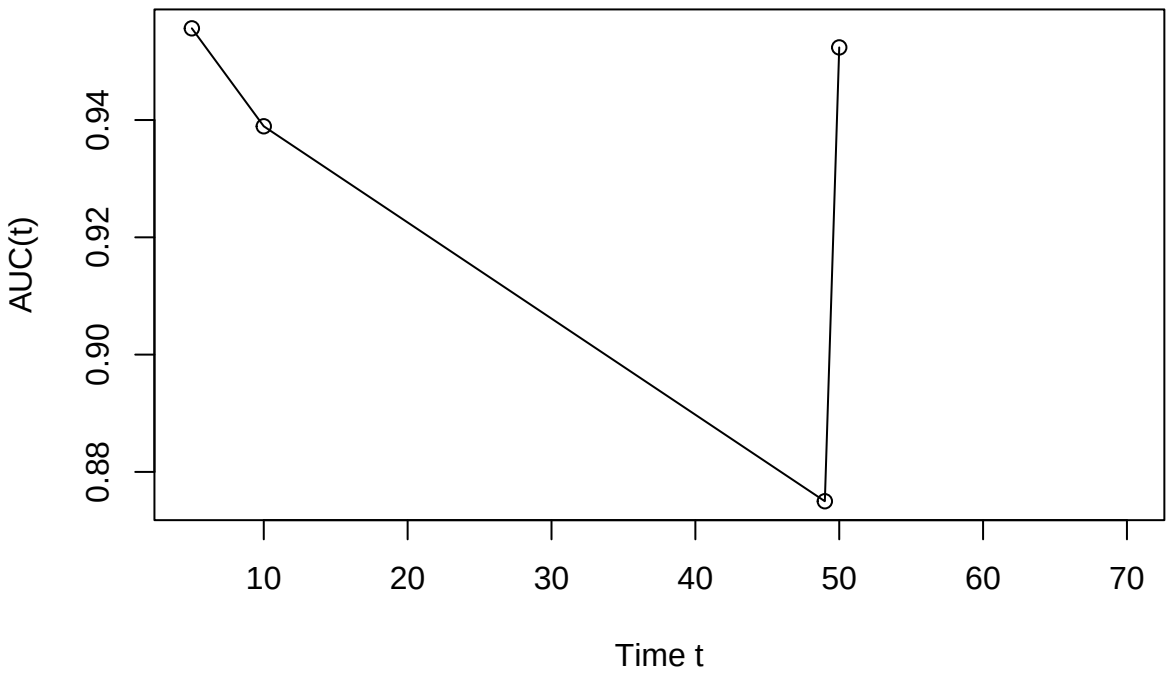
coeffient plot

```
coef_plot_DMS_P_7 <- ggforest(cox_DMS_P_7_heav , data = DMS_P_7_cut,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_DMS_P_7
```



Area under the curve

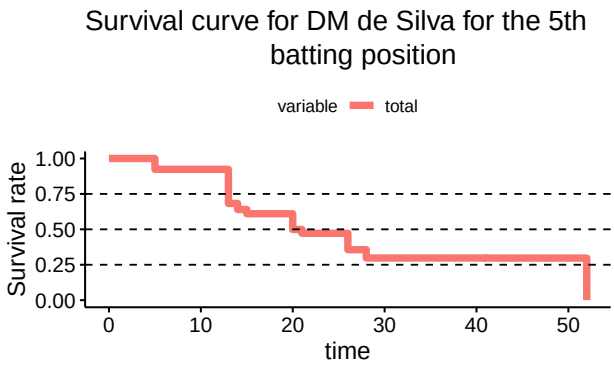
```
AUC(Surv(start,cum_balls , wicket) ~
      cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = DMS_P_7_cut
```

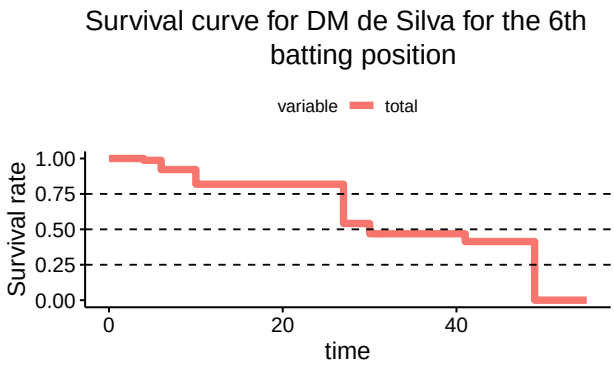


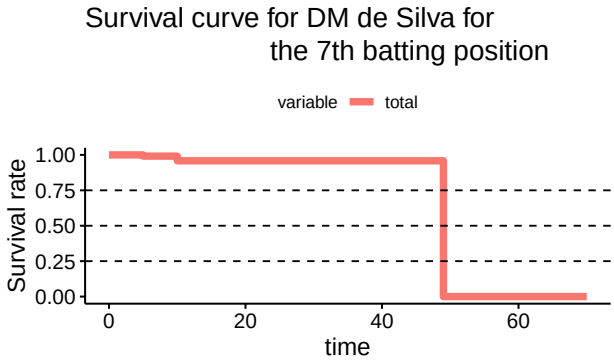

```
## $AUCt
##   time       AUC
## 1    5 0.9556314
## 2   10 0.9389313
## 3   49 0.8750000
## 4   50 0.9523810
## 5   70       NaN
##
## $AUC
## [1] 0.945
```

Survival curve

```
DMSSurvPlotPos <- ggarrange(cox_plot_DMS_P_5,cox_plot_DMS_P_6,cox_plot_DMS_P_7)
DMSSurvPlotPos
```







Upload to texStudio

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

Aalen model

```
DMS_p5 <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
      cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = DMS_P_5_cut)

DMS_p6 <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
      cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = DMS_P_6_cut)

DMS_p7 <- autoplot(aareg(Surv(start,cum_balls , wicket) ~
      cum_singles +hv1 + hv2 + cum_boundaries + ridge(cum_dotBalls) + cluster (id) , data = DMS_P_7_cut)

DMSAalenPos <- ggarrange(DMS_p5,DMS_p6,DMS_p7, nrow = 3)
DMSAalenPos
```


cluster(id)
cum_boundaries
cum_singles
hv1
hv2
Intercept
ridge(cum_dotBalls)
variable
cluster(id)
cum_boundaries
cum_singles
hv1
hv2
Intercept
ridge(cum_dotBalls)

```
filepath <- "~/Research/Thesis_S16373/images/"
postscript(file = paste0(filepath,"DMSAalenPos.eps"),width = 10,height = 20,horizontal = FALSE)
DMSAalenPos
dev.off()
```