MDKJ_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 MD Gunathilaka

```
bt10 <- data_merge %>% filter(grepl('MDKJ Perera' , Striker))
```

1)Analysis for two different innings

First innings

Filter the first innings

```
MDKJ_1 <- bt10 %>% filter(innings == 1)
```

Cox model

```
cox_MDKJ_1_new <- coxph(Surv(cum_balls , wicket) ~
                        cum_runs + cum_singles + ridge(cum_dotBalls), data = MDKJ_1)

summary(cox_MDKJ_1_new)
```

```
## Call:
## coxph(formula = Surv(cum_balls, wicket) ~ cum_runs + cum_singles +
##     ridge(cum_dotBalls), data = MDKJ_1)
##
##   n= 1177, number of events= 30
##
##                      coef    se(coef) se2      Chisq DF p
## cum_runs            -0.2197 0.03493  0.03387 39.55 1  3.2e-10
## cum_singles         -0.3003 0.07028  0.06217 18.26 1  1.9e-05
## ridge(cum_dotBalls) -0.3942 0.07319  0.05043 29.01 1  7.2e-08
##
##                      exp(coef) exp(-coef) lower .95 upper .95
## cum_runs               0.8028     1.246    0.7496    0.8596
## cum_singles            0.7406     1.350    0.6453    0.8499
## ridge(cum_dotBalls)    0.6742     1.483    0.5841    0.7782
##
## Iterations: 3 outer, 14 Newton-Raphson
## Degrees of freedom for terms= 0.9 0.8 0.5
## Concordance= 0.982  (se = 0.004 )
## Likelihood ratio test= 158.8  on 2.2 df,   p=<2e-16
```

Test pH assumptions

```
test_MDKJ_1 <- cox.zph(cox_MDKJ_1_new)
test_MDKJ_1
```
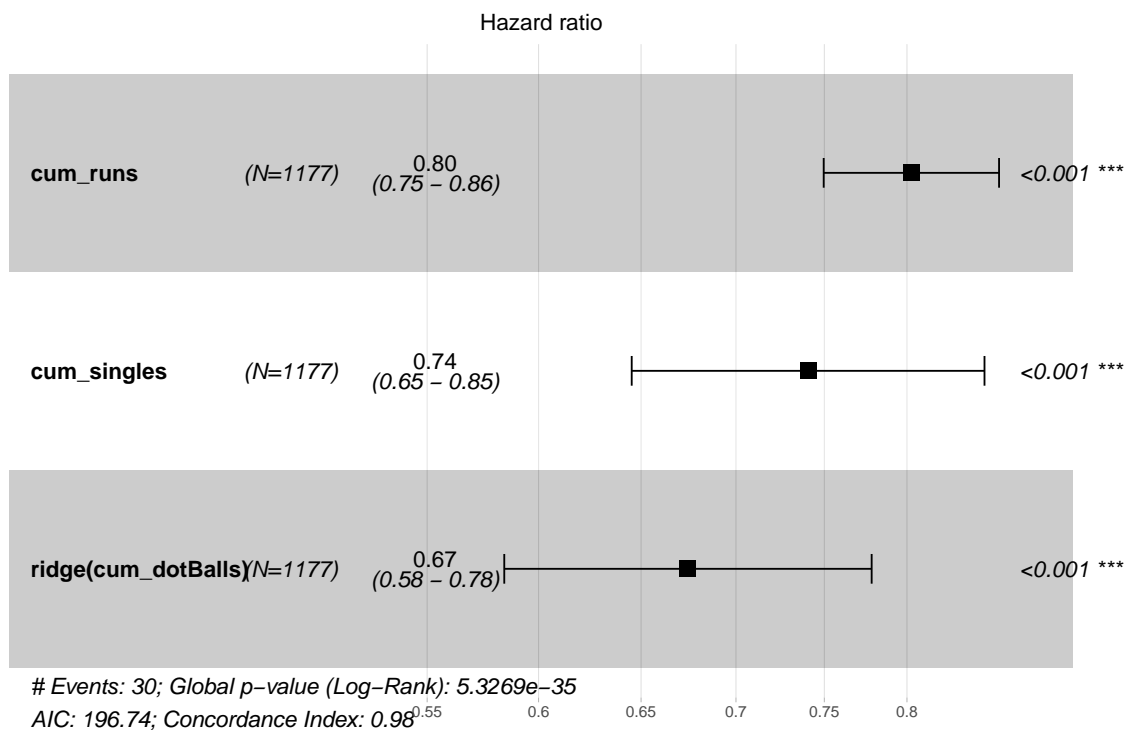
```
##                       chisq  df    p
## cum_runs            1.30e-01 0.94 0.69
## cum_singles         2.09e+00 0.78 0.11
## ridge(cum_dotBalls) 4.31e-05 0.47 0.91
## GLOBAL              1.36e+00 2.20 0.55
```

Plot the survival curve

```
cox_plot_MDKJ_1 <- ggadjustedcurves(cox_MDKJ_1_new , data = MDKJ_1 , size = 2,ggtheme = theme_survminer
                    geom_hline(yintercept = c(0.25,0.5,0.75) , linetype = "dashed")  +
                        ggtitle("Survival curve for MDKJ Perera for the first innings") +
    theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
```

coeffient plot

```
coef_plot_MDKJ_1 <- ggforest(cox_MDKJ_1_new , data = MDKJ_1,fontsize = 1.2) +
    theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_MDKJ_1
```

Hazard ratio

cum_runs	*(N=1177)*	0.80 *(0.75 – 0.86)*	■	*<0.001 ****
cum_singles	*(N=1177)*	0.74 *(0.65 – 0.85)*	■	*<0.001 ****
ridge(cum_dotBalls)	*(N=1177)*	0.67 *(0.58 – 0.78)*	■	*<0.001 ****

 # Events: 30; Global p–value (Log–Rank): 5.3269e–35
AIC: 196.74; Concordance Index: 0.98   0.55   0.6   0.65   0.7   0.75   0.8

Area under the curve

```
AUC(Surv(cum_balls , wicket) ~
        cum_runs + cum_singles + ridge(cum_dotBalls), data = MDKJ_1)
```
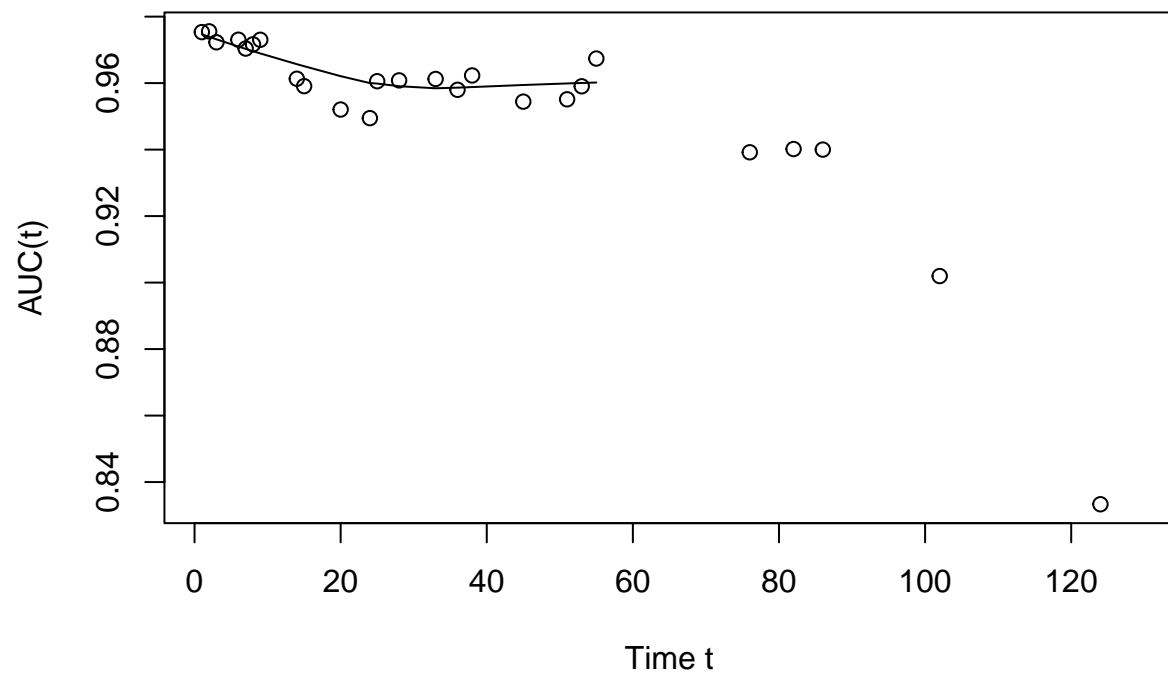
```
## $AUCt
##     time       AUC
## 1      1 0.9753401
## 2      2 0.9755672
## 3      3 0.9722719
## 4      6 0.9730510
## 5      7 0.9703850
## 6      8 0.9716599
## 7      9 0.9730290
## 8     14 0.9613130
## 9     15 0.9590854
## 10    20 0.9520548
## 11    24 0.9494640
## 12    25 0.9605678
## 13    28 0.9608177
## 14    33 0.9612403
## 15    36 0.9579832
## 16    38 0.9623060
## 17    45 0.9544236
## 18    51 0.9551282
## 19    53 0.9590444
## 20    55 0.9673913
## 21    76 0.9391892
## 22    82 0.9401709
## 23    86 0.9400000
## 24   102 0.9019608
```

```
## 25   124 0.8333333
## 26   129        NaN
##
## $AUC
## [1] 0.9649855
```

Second Innings

Filter the second innings

```r
MDKJ_2 <- bt10 %>% filter(innings == 2)
```

Cox model

```r
cox_MDKJ_2_new <- coxph(Surv(cum_balls , wicket) ~
                        cum_runs + cum_singles + ridge(cum_dotBalls), data = MDKJ_2)
summary(cox_MDKJ_2_new)
```

```
## Call:
## coxph(formula = Surv(cum_balls, wicket) ~ cum_runs + cum_singles +
##     ridge(cum_dotBalls), data = MDKJ_2)
##
##    n= 793, number of events= 21
##
##                       coef    se(coef)  se2     Chisq DF p
## cum_runs            -0.1424  0.03259   0.0301  19.09 1  1.2e-05
## cum_singles         -0.5014  0.10981   0.1084  20.85 1  5.0e-06
## ridge(cum_dotBalls) -0.2584  0.07134   0.0453  13.12 1  2.9e-04
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## cum_runs               0.8673      1.153    0.8136    0.9245
## cum_singles            0.6057      1.651    0.4884    0.7511
## ridge(cum_dotBalls)    0.7723      1.295    0.6715    0.8882
##
## Iterations: 2 outer, 10 Newton-Raphson
## Degrees of freedom for terms= 0.9 1.0 0.4
## Concordance= 0.96  (se = 0.011 )
## Likelihood ratio test= 90.45  on 2.23 df,   p=<2e-16
```

Test pH assumptions

```r
test_MDKJ_2 <- cox.zph(cox_MDKJ_2_new)
test_MDKJ_2
```

5
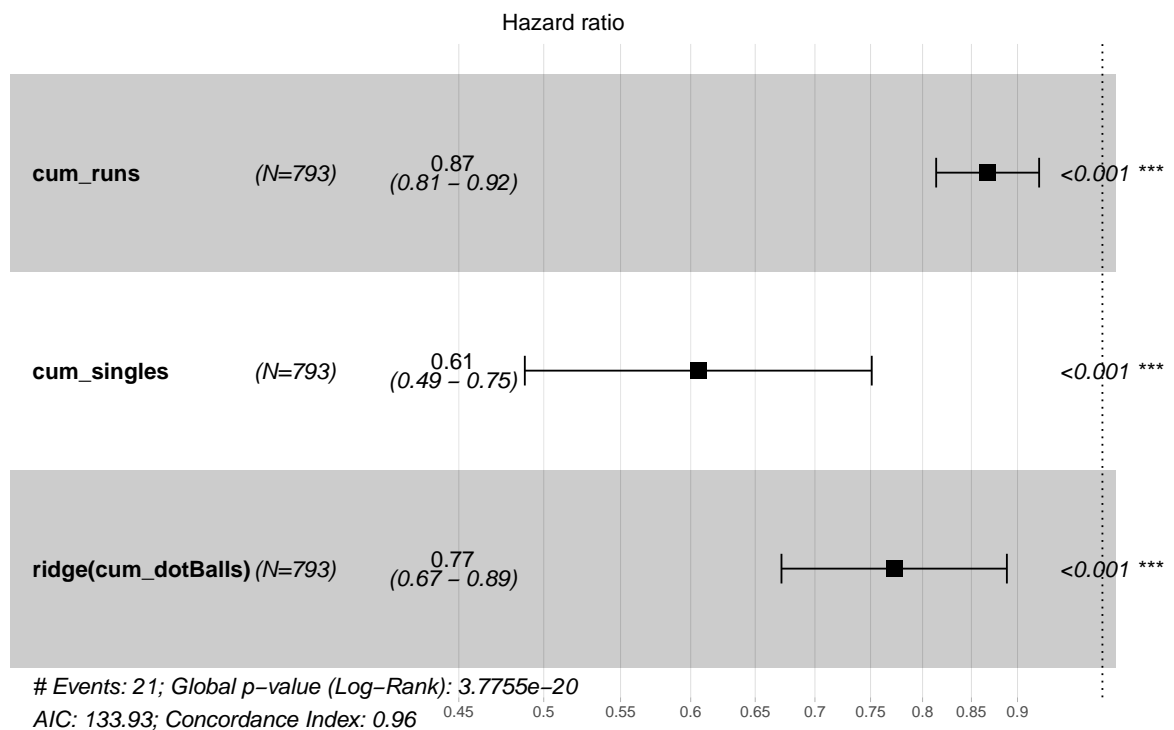
```
##                        chisq   df    p
## cum_runs             2.89133 0.85 0.072
## cum_singles          2.26465 0.98 0.128
## ridge(cum_dotBalls)  0.00378 0.40 0.692
## GLOBAL               2.85767 2.23 0.279
```

Plot the survival curve

```
cox_plot_MDKJ_2 <- ggadjustedcurves(cox_MDKJ_2_new , data = MDKJ_2 , size = 2,ggtheme = theme_survminer
                   geom_hline(yintercept = c(0.25,0.5,0.75) , linetype = "dashed")  +
                      ggtitle("Survival curve for MDKJ Perera for the second innings") +
    theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
```

coeffient plot

```
coef_plot_MDKJ_2 <- ggforest(cox_MDKJ_2_new , data = MDKJ_2,fontsize = 1.2) +
  theme(axis.title = element_text(size = 15),axis.text = element_text(size = 14))
coef_plot_MDKJ_2
```

Hazard ratio

cum_runs	*(N=793)*	0.87 (0.81 – 0.92)	▪	<0.001 ***
cum_singles	*(N=793)*	0.61 (0.49 – 0.75)	▪	<0.001 ***
ridge(cum_dotBalls)	*(N=793)*	0.77 (0.67 – 0.89)	▪	<0.001 ***

 # Events: 21; Global p−value (Log−Rank): 3.7755e−20
AIC: 133.93; Concordance Index: 0.96

0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9

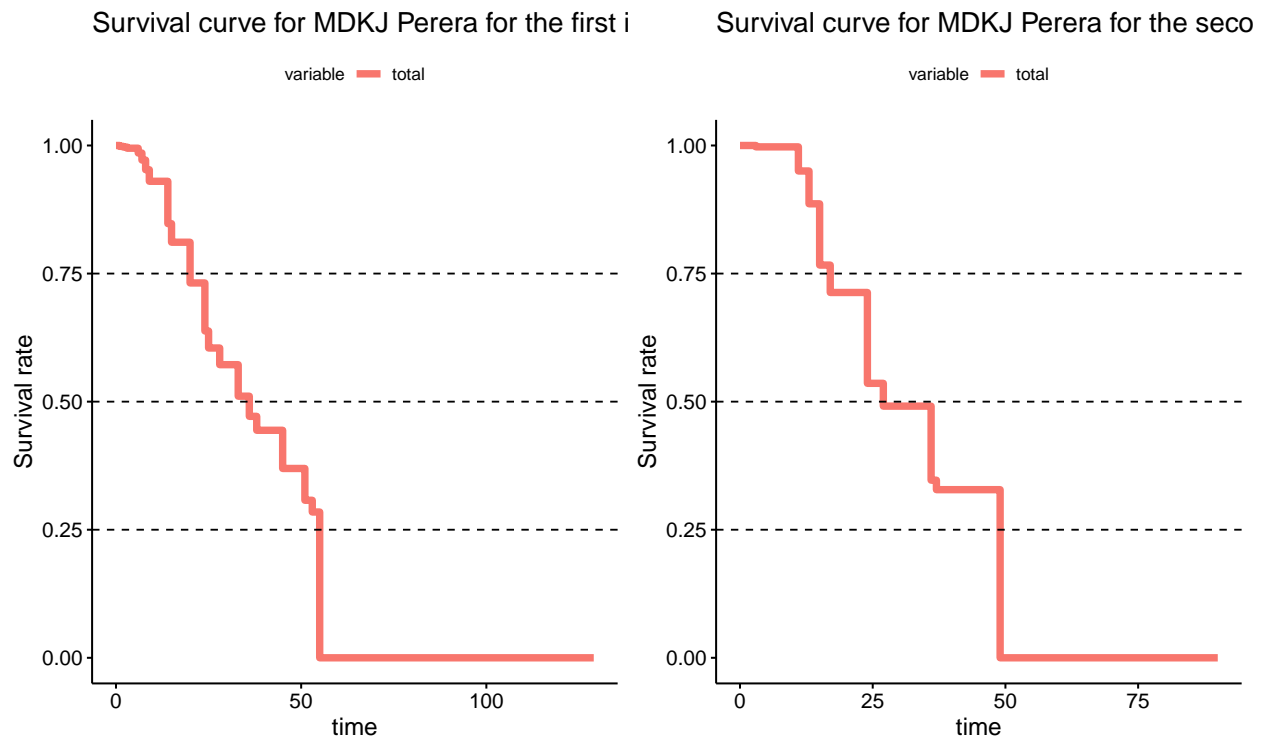**Area under the curve**

```
AUC(Surv(cum_balls , wicket) ~
      cum_runs + cum_singles + ridge(cum_dotBalls), data = MDKJ_2)
```
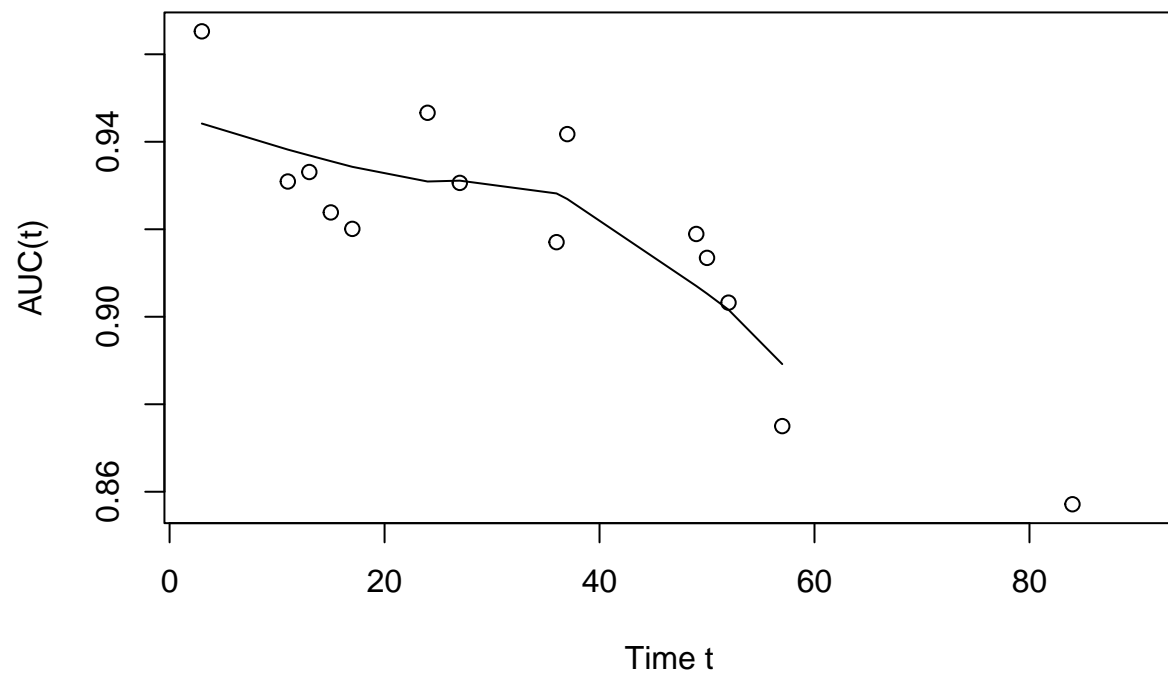
```
## $AUCt
##    time       AUC
## 1     3 0.9652406
## 2    11 0.9309154
## 3    13 0.9330855
## 4    15 0.9238477
## 5    17 0.9200864
## 6    24 0.9466292
## 7    27 0.9305994
## 8    36 0.9170507
## 9    37 0.9417476
## 10   49 0.9189189
## 11   50 0.9134615
## 12   52 0.9032258
## 13   57 0.8750000
## 14   84 0.8571429
## 15   90       NaN
##
## $AUC
## [1] 0.9338747
```

Survival curve

```
MDKJSurvPlot <- ggarrange(cox_plot_MDKJ_1,cox_plot_MDKJ_2,ncol = 2)
MDKJSurvPlot
```
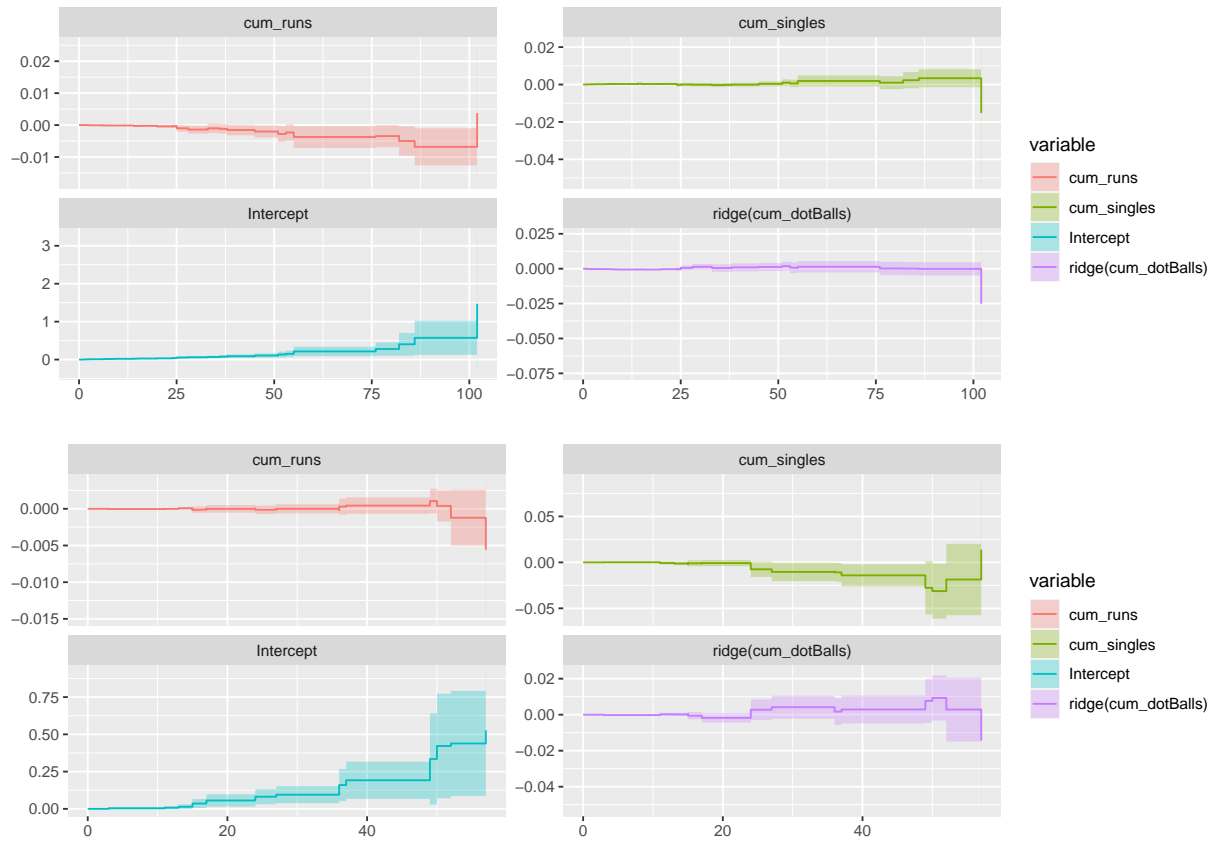

Upload to texStudio

```
filepath <- "~/Research/Thesis_S16373/images/"
postscript(file = paste0(filepath,"MDKJSurvPlot.eps"),width = 10,height = 6,horizontal = FALSE)
MDKJSurvPlot
dev.off()
```

Aalen model

```
MDKJ_p1 <- autoplot(aareg(Surv(cum_balls , wicket) ~
                          cum_runs + cum_singles + ridge(cum_dotBalls), data = MDKJ_1))

MDKJ_p2 <- autoplot(aareg(Surv(cum_balls , wicket) ~
                          cum_runs + cum_singles + ridge(cum_dotBalls), data = MDKJ_2))

MDKJAalen <- ggarrange(MDKJ_p1,MDKJ_p2, nrow = 2)
MDKJAalen
```

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