project2

Ze Li

2023-12-18

Libraries

```
library(tidyverse)
library(readr)
library(MASS)
library(car)
library(dplyr)
library(leaps)
library(survival)
library(survminer)
```

Data Clean

```
breastcancer_data =
  read_csv("Project_2_data.csv") |>
  janitor::clean_names()
```

```
## Rows: 4024 Columns: 16
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (11): Race, Marital Status, T Stage, N Stage, 6th Stage, differentiate, ...
## dbl  (5): Age, Tumor Size, Regional Node Examined, Reginol Node Positive, Su...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
bc = breastcancer_data |>
  mutate(
    race = factor(race, levels = c("White", "Black", "Other")),
    marital_status = factor(marital_status, levels = c("Married", "Divorced",
                                                       "Single", "Widowed",
                                                       "Separated")),
    t_stage = factor(t_stage, levels = c("T1", "T2", "T3", "T4")),
    n_stage = factor(n_stage, levels = c("N1", "N2", "N3")),
    x6th_stage = factor(x6th_stage, levels = c("IIA", "IIIA", "IIIC", "IIB", "IIIB")),
    differentiate = factor(differentiate, levels = c("Poorly differentiated",
                                                     "Moderately differentiated",
```

```
                                                    "Well differentiated",
                                                    "Undifferentiated")),
    grade = factor(grade, levels = c("1", "2", "3", "anaplastic; Grade IV")),
    a_stage = factor(a_stage, levels = c("Distant", "Regional")),
    estrogen_status = factor(estrogen_status, levels = c("Negative", "Positive")),
    progesterone_status = factor(progesterone_status,
                                 levels = c("Negative", "Positive")),
    status = factor(status, levels = c("Dead", "Alive"))
  )
```

Descriptive statistics for all variables

```
summary(bc)
```

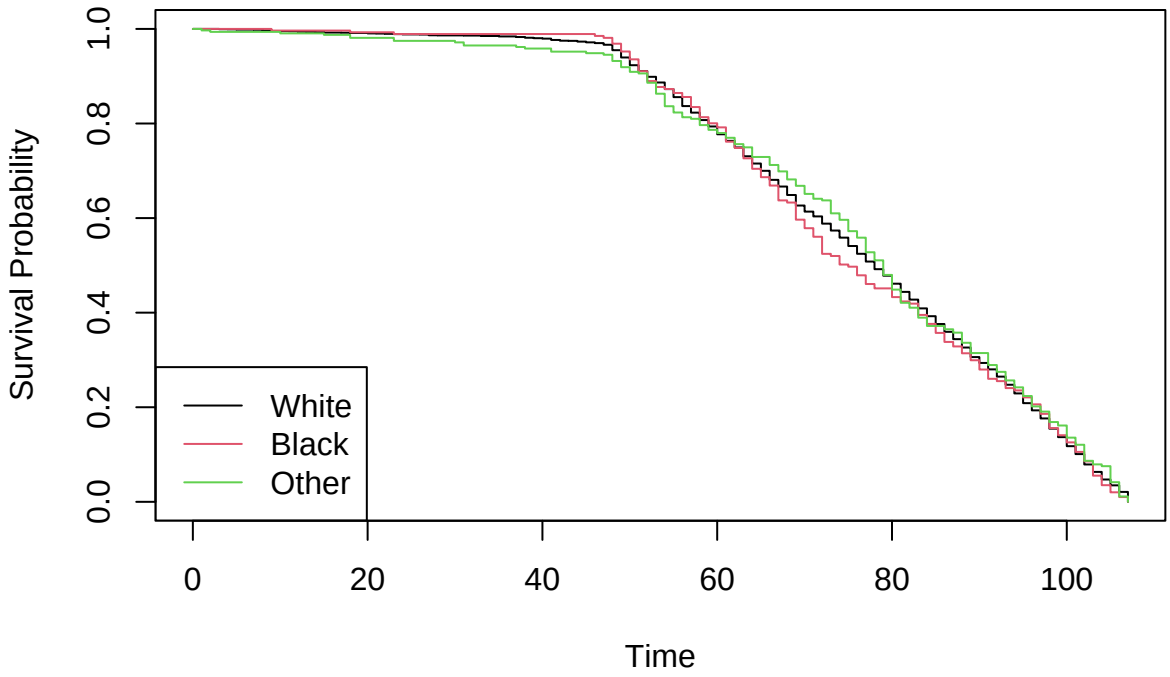
```
##       age           race        marital_status t_stage   n_stage   x6th_stage
##  Min.   :30.00   White:3413   Married  :2643   T1:1603   N1:2732   IIA :1305
##  1st Qu.:47.00   Black: 291   Divorced : 486   T2:1786   N2: 820   IIIA:1050
##  Median :54.00   Other: 320   Single   : 615   T3: 533   N3: 472   IIIC: 472
##  Mean   :53.97                Widowed  : 235   T4: 102             IIB :1130
##  3rd Qu.:61.00                Separated:  45                       IIIB:  67
##  Max.   :69.00
##                    differentiate                    grade          a_stage
##  Poorly differentiated    :1111   1                   : 543   Distant :  92
##  Moderately differentiated:2351   2                   :2351   Regional:3932
##  Well differentiated      : 543   3                   :1111
##  Undifferentiated         :  19   anaplastic; Grade IV:  19
##
##
##    tumor_size     estrogen_status progesterone_status regional_node_examined
##  Min.   :  1.00   Negative: 269   Negative: 698       Min.   : 1.00
##  1st Qu.: 16.00   Positive:3755   Positive:3326       1st Qu.: 9.00
##  Median : 25.00                                       Median :14.00
##  Mean   : 30.47                                       Mean   :14.36
##  3rd Qu.: 38.00                                       3rd Qu.:19.00
##  Max.   :140.00                                       Max.   :61.00
##  reginol_node_positive survival_months   status
##  Min.   : 1.000        Min.   :  1.0   Dead : 616
##  1st Qu.: 1.000        1st Qu.: 56.0   Alive:3408
##  Median : 2.000        Median : 73.0
##  Mean   : 4.158        Mean   : 71.3
##  3rd Qu.: 5.000        3rd Qu.: 90.0
##  Max.   :46.000        Max.   :107.0
```

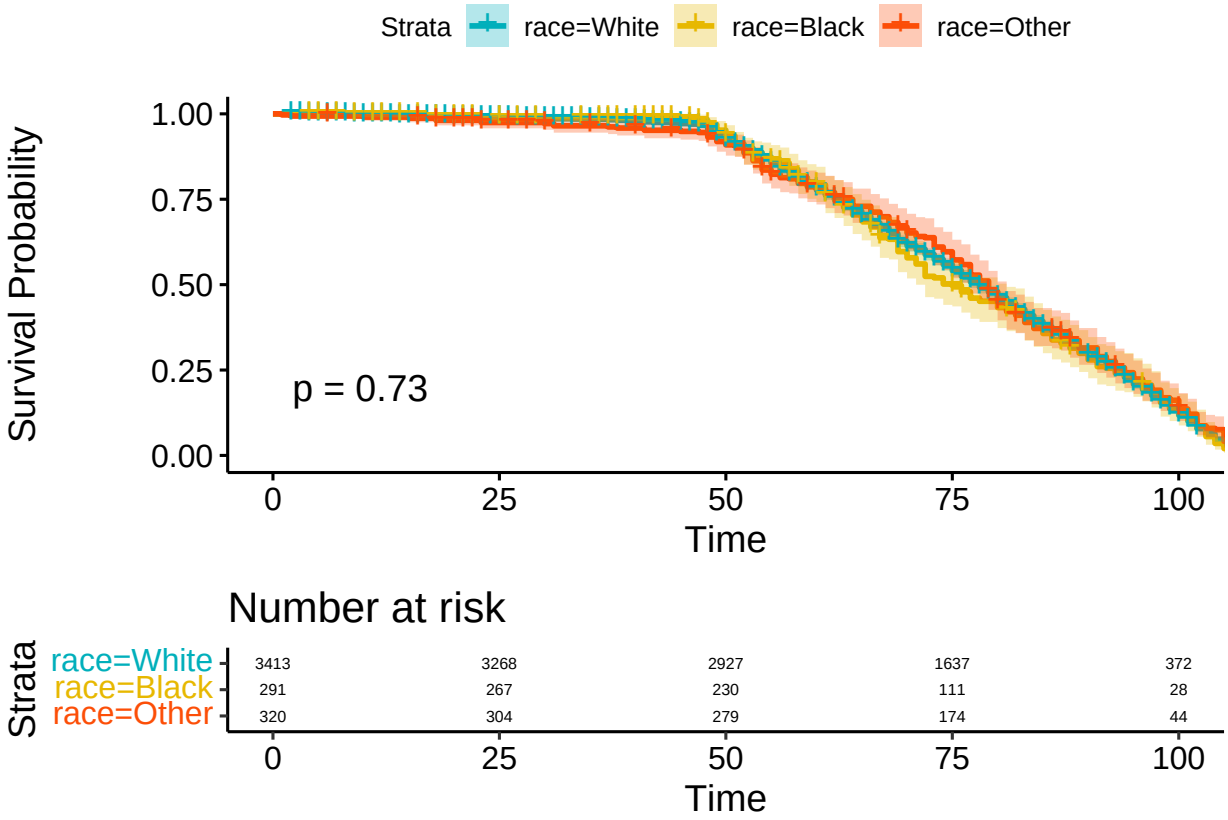
Kaplan-Meier survival curves

```
bc$survival_months <- as.numeric(bc$survival_months)
bc$status <- as.numeric(bc$status)
surv_obj <- Surv(time = bc$survival_months, event = bc$status)
```

```r
km_fit <- survfit(surv_obj ~ race, data = bc)
plot(km_fit, col = 1:3, xlab = "Time", ylab = "Survival Probability")
legend("bottomleft", legend = levels(bc$race), col = 1:3, lty = 1)

ggsurvplot(km_fit, data = bc, pval = TRUE, conf.int = TRUE,
           xlab = "Time", ylab = "Survival Probability",
           risk.table = TRUE, # Add risk table
           risk.table.height = 0.3, # Adjust the height of the risk table
           risk.table.fontsize = 2, # Adjust the font size in the risk table
           conf.int.size = 0.5, # Adjust the size of the confidence interval lines
           palette = c("#00AFBB", "#E7B800", "#FC4E07"))
```





The Kaplan-Meier plot displays survival probabilities for three racial groups: White, Black, and Other, over time. Survival curves are closely aligned, suggesting minimal differences in survival outcomes across these groups, which is statistically supported by a p-value of 0.73, indicating no significant difference. The number-at-risk table below the plot shows a decreasing count over time, reflecting those still under observation at specific time points, with the White group starting with the highest count.

Therefore, the consistency of the curves along with the non-significant p-value suggests race, in this sample, does not have a differential impact on survival probability.

Fit a Cox Proportional Hazards Model

```
bc$survival_months <- as.numeric(bc$survival_months)
bc$status <- as.numeric(bc$status)
surv_obj <- Surv(time = bc$survival_months, event = bc$status)

# Fit the Cox model
cox_model <- coxph(surv_obj ~ age + race + marital_status + t_stage + n_stage + x6th_stage +
                     differentiate + grade + a_stage + tumor_size + estrogen_status +
                     progesterone_status + regional_node_examined + reginol_node_positive, data = bc)
summary(cox_model)
```

```
## Call:
## coxph(formula = surv_obj ~ age + race + marital_status + t_stage +
##     n_stage + x6th_stage + differentiate + grade + a_stage +
```

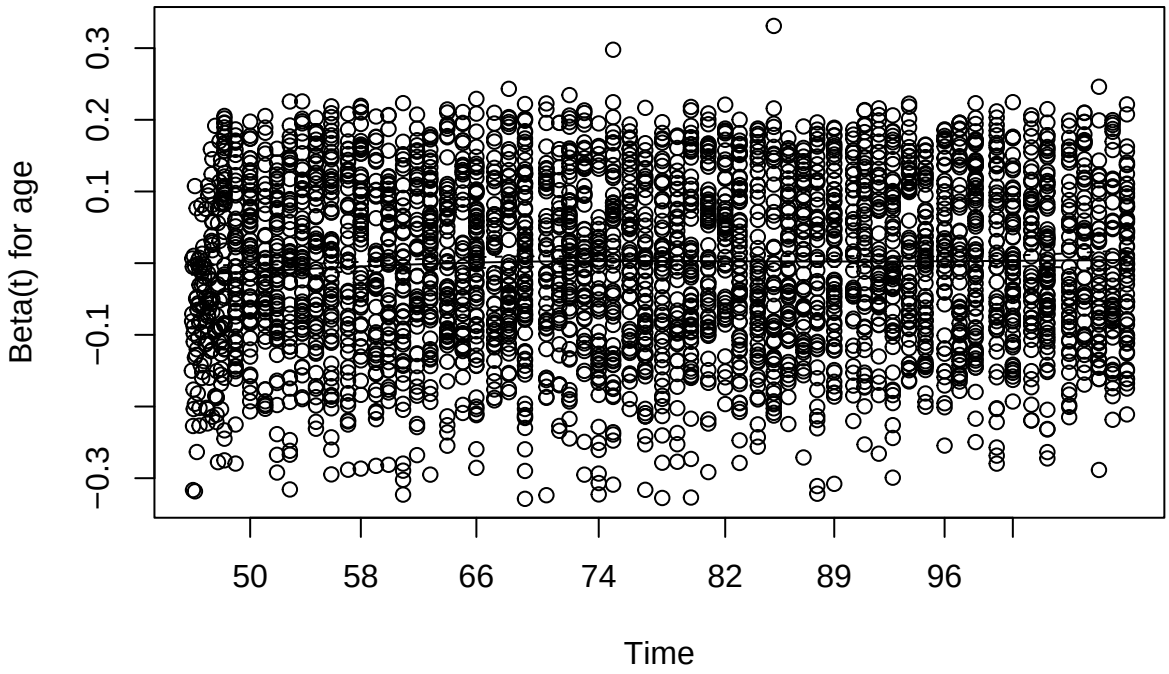
```
##     tumor_size + estrogen_status + progesterone_status + regional_node_examined +
##     reginol_node_positive, data = bc)
##
##   n= 4024, number of events= 3408
##
##                                          coef  exp(coef)   se(coef)      z
## age                                   -0.0004124  0.9995877  0.0020594 -0.200
## raceBlack                              0.0725419  1.0752378  0.0713977  1.016
## raceOther                             -0.0282799  0.9721162  0.0623962 -0.453
## marital_statusDivorced                -0.0604365  0.9413536  0.0549382 -1.100
## marital_statusSingle                  -0.0482234  0.9529209  0.0496924 -0.970
## marital_statusWidowed                 -0.0303772  0.9700795  0.0782217 -0.388
## marital_statusSeparated                0.2309519  1.2597987  0.1848733  1.249
## t_stageT2                              0.0125409  1.0126199  0.0857679  0.146
## t_stageT3                             -0.0847136  0.9187754  0.1434176 -0.591
## t_stageT4                             -0.3011978  0.7399314  0.2924931 -1.030
## n_stageN2                             -0.1666794  0.8464709  0.1012458 -1.646
## n_stageN3                             -0.0261531  0.9741859  0.1405331 -0.186
## x6th_stageIIIA                         0.1632452  1.1773253  0.1182755  1.380
## x6th_stageIIIC                                NA         NA  0.0000000     NA
## x6th_stageIIB                         -0.0071028  0.9929224  0.0905328 -0.078
## x6th_stageIIIB                         0.1569403  1.1699257  0.3163929  0.496
## differentiateModerately differentiated 0.0557485  1.0573318  0.0419336  1.329
## differentiateWell differentiated       0.1215000  1.1291894  0.0580404  2.093
## differentiateUndifferentiated         -0.3997794  0.6704679  0.3199544 -1.249
## grade2                                        NA         NA  0.0000000     NA
## grade3                                        NA         NA  0.0000000     NA
## gradeanaplastic; Grade IV                     NA         NA  0.0000000     NA
## a_stageRegional                       -0.0051530  0.9948603  0.1508261 -0.034
## tumor_size                             0.0009129  1.0009133  0.0018429  0.495
## estrogen_statusPositive                0.0915971  1.0959232  0.0909893  1.007
## progesterone_statusPositive            0.1628328  1.1768400  0.0545896  2.983
## regional_node_examined                 0.0025479  1.0025511  0.0023305  1.093
## reginol_node_positive                 -0.0032452  0.9967601  0.0080101 -0.405
##                                       Pr(>|z|)
## age                                    0.84127
## raceBlack                              0.30962
## raceOther                              0.65038
## marital_statusDivorced                 0.27130
## marital_statusSingle                   0.33183
## marital_statusWidowed                  0.69776
## marital_statusSeparated                0.21158
## t_stageT2                              0.88375
## t_stageT3                              0.55474
## t_stageT4                              0.30312
## n_stageN2                              0.09970 .
## n_stageN3                              0.85237
## x6th_stageIIIA                         0.16752
## x6th_stageIIIC                              NA
## x6th_stageIIB                          0.93747
## x6th_stageIIIB                         0.61987
## differentiateModerately differentiated 0.18370
## differentiateWell differentiated       0.03632 *
## differentiateUndifferentiated          0.21149
```

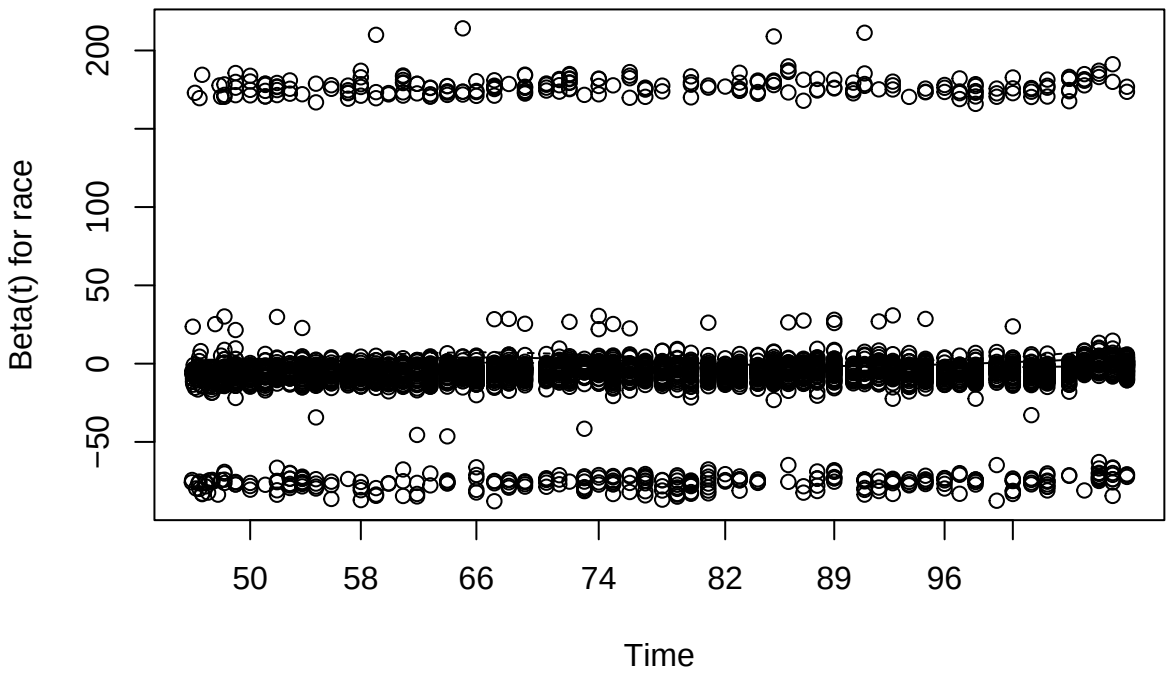
```
## grade2                                        NA
## grade3                                        NA
## gradeanaplastic; Grade IV                     NA
## a_stageRegional                            0.97275
## tumor_size                                 0.62035
## estrogen_statusPositive                    0.31409
## progesterone_statusPositive                0.00286 **
## regional_node_examined                     0.27428
## reginol_node_positive                      0.68538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                         exp(coef) exp(-coef) lower .95 upper .95
## age                                        0.9996     1.0004    0.9956     1.004
## raceBlack                                  1.0752     0.9300    0.9348     1.237
## raceOther                                  0.9721     1.0287    0.8602     1.099
## marital_statusDivorced                     0.9414     1.0623    0.8453     1.048
## marital_statusSingle                       0.9529     1.0494    0.8645     1.050
## marital_statusWidowed                      0.9701     1.0308    0.8322     1.131
## marital_statusSeparated                    1.2598     0.7938    0.8769     1.810
## t_stageT2                                  1.0126     0.9875    0.8559     1.198
## t_stageT3                                  0.9188     1.0884    0.6936     1.217
## t_stageT4                                  0.7399     1.3515    0.4171     1.313
## n_stageN2                                  0.8465     1.1814    0.6941     1.032
## n_stageN3                                  0.9742     1.0265    0.7396     1.283
## x6th_stageIIIA                             1.1773     0.8494    0.9337     1.484
## x6th_stageIIIC                                 NA         NA        NA        NA
## x6th_stageIIB                              0.9929     1.0071    0.8315     1.186
## x6th_stageIIIB                             1.1699     0.8548    0.6293     2.175
## differentiateModerately differentiated     1.0573     0.9458    0.9739     1.148
## differentiateWell differentiated           1.1292     0.8856    1.0078     1.265
## differentiateUndifferentiated              0.6705     1.4915    0.3581     1.255
## grade2                                         NA         NA        NA        NA
## grade3                                         NA         NA        NA        NA
## gradeanaplastic; Grade IV                      NA         NA        NA        NA
## a_stageRegional                            0.9949     1.0052    0.7403     1.337
## tumor_size                                 1.0009     0.9991    0.9973     1.005
## estrogen_statusPositive                    1.0959     0.9125    0.9169     1.310
## progesterone_statusPositive                1.1768     0.8497    1.0574     1.310
## regional_node_examined                     1.0026     0.9975    0.9980     1.007
## reginol_node_positive                      0.9968     1.0033    0.9812     1.013
##
## Concordance= 0.538  (se = 0.006 )
## Likelihood ratio test= 39.61  on 24 df,   p=0.02
## Wald test            = 38.29  on 24 df,   p=0.03
## Score (logrank) test = 38.42  on 24 df,   p=0.03
```

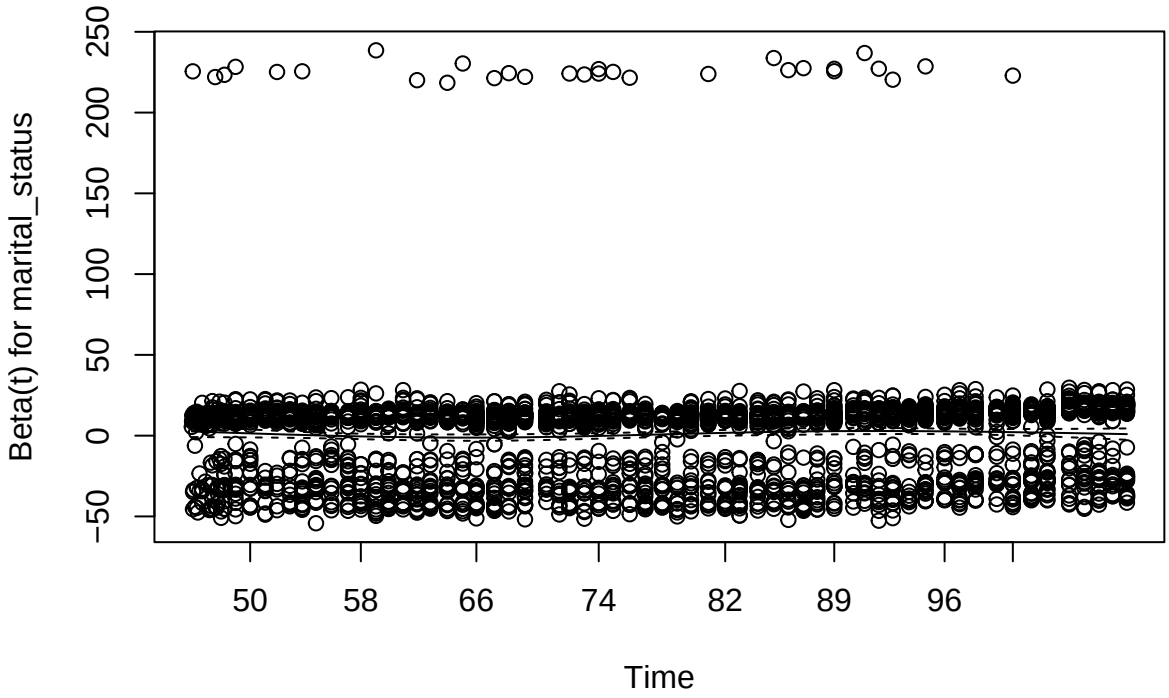
Concordance= 0.538 (se = 0.006)

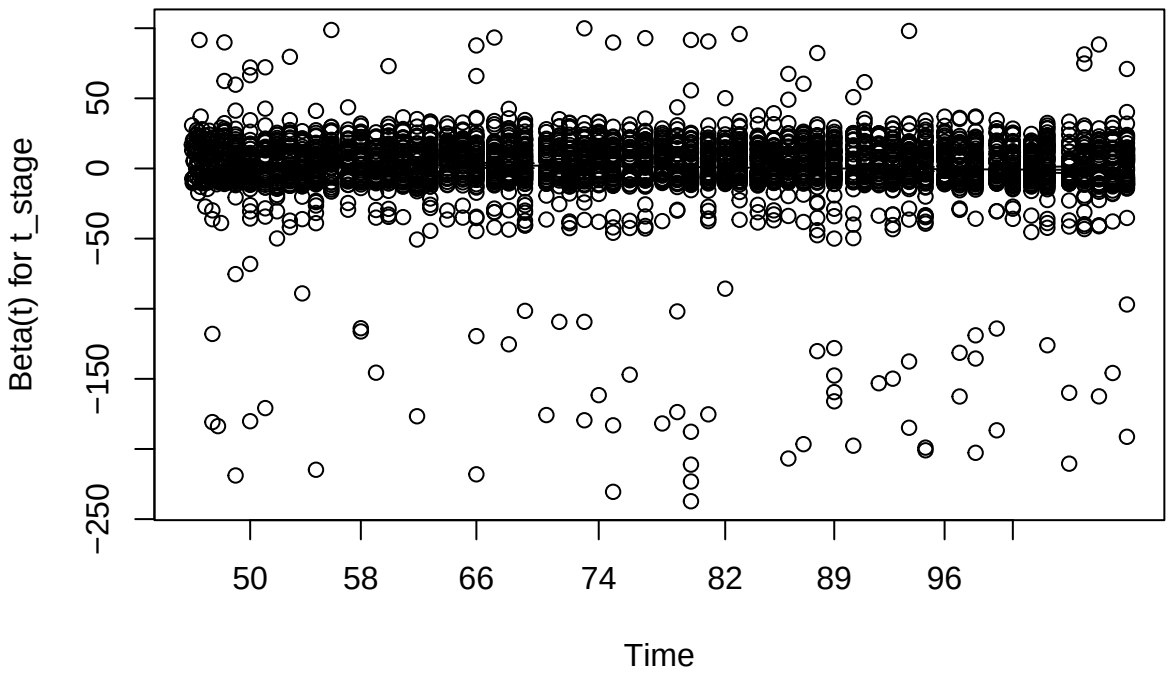
Check Proportional Hazards Assumption

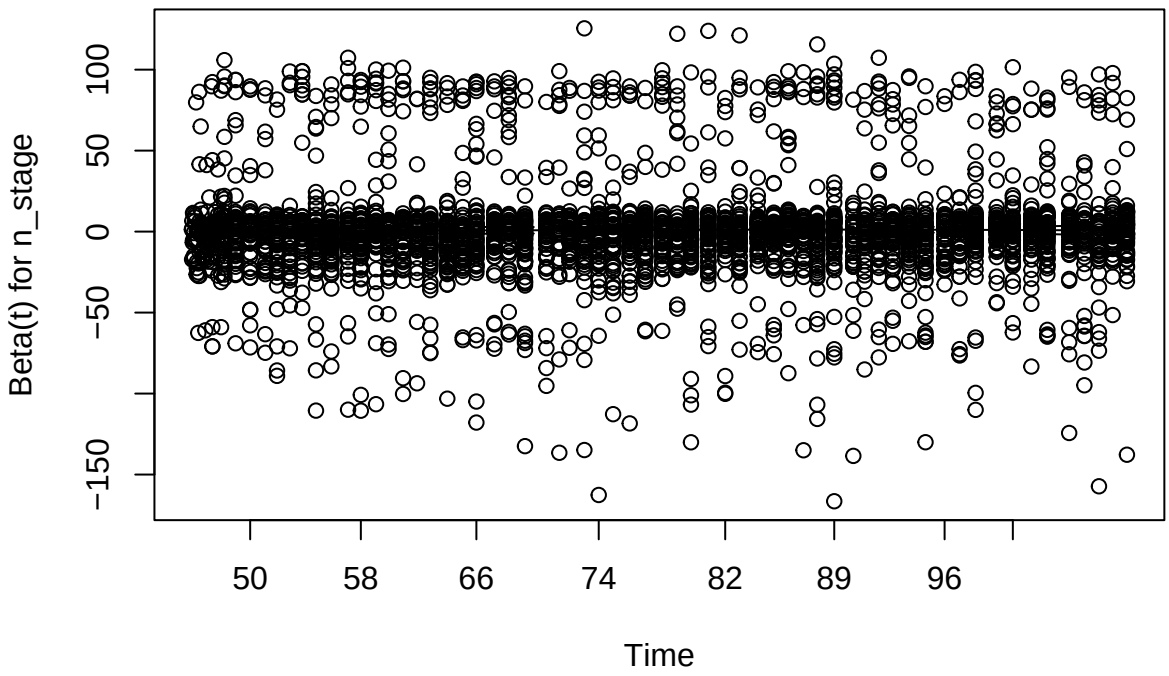
```r
zph <- cox.zph(cox_model)
plot(zph)
```

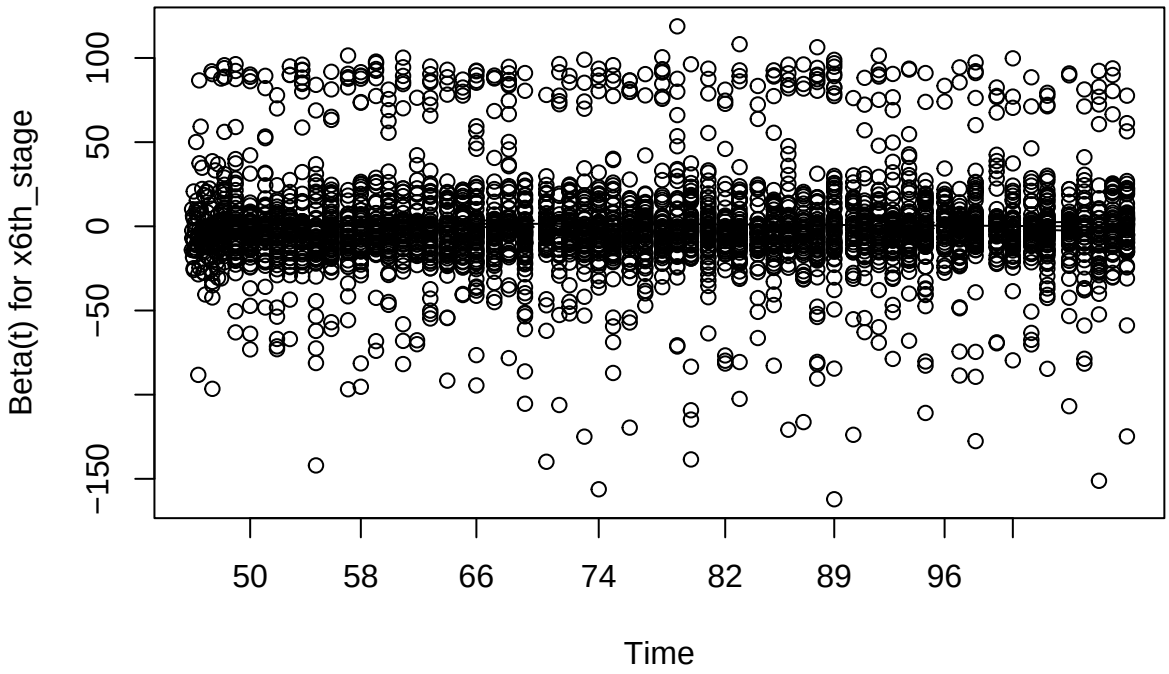


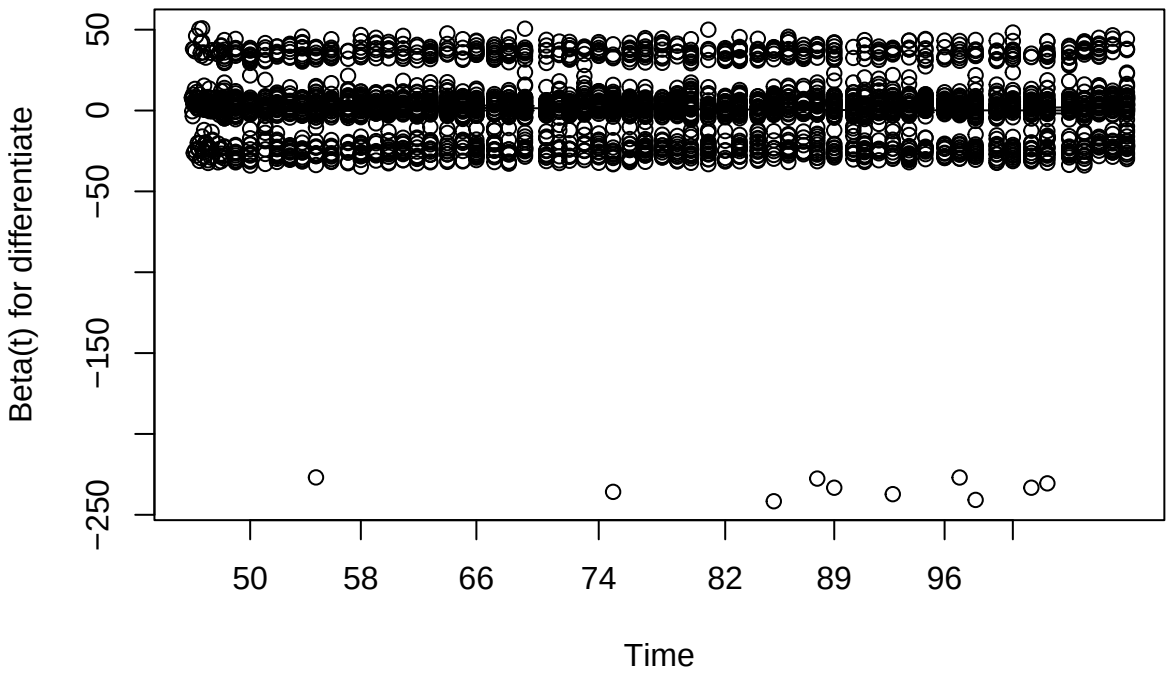

Beta(t) for race
200
100
50
0
−50
50
58
66
74
82
89
96
Time

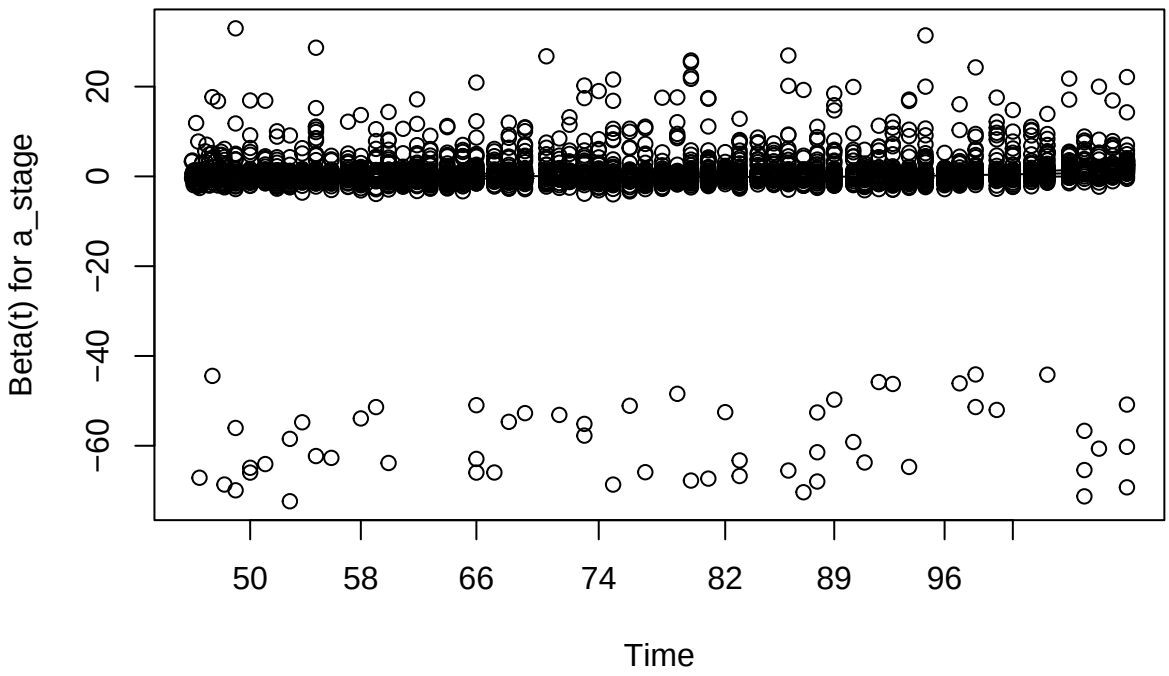

Beta(t) for marital_status
250
200
150
100
50
0
−50
50
58
66
74
82
89
96
Time

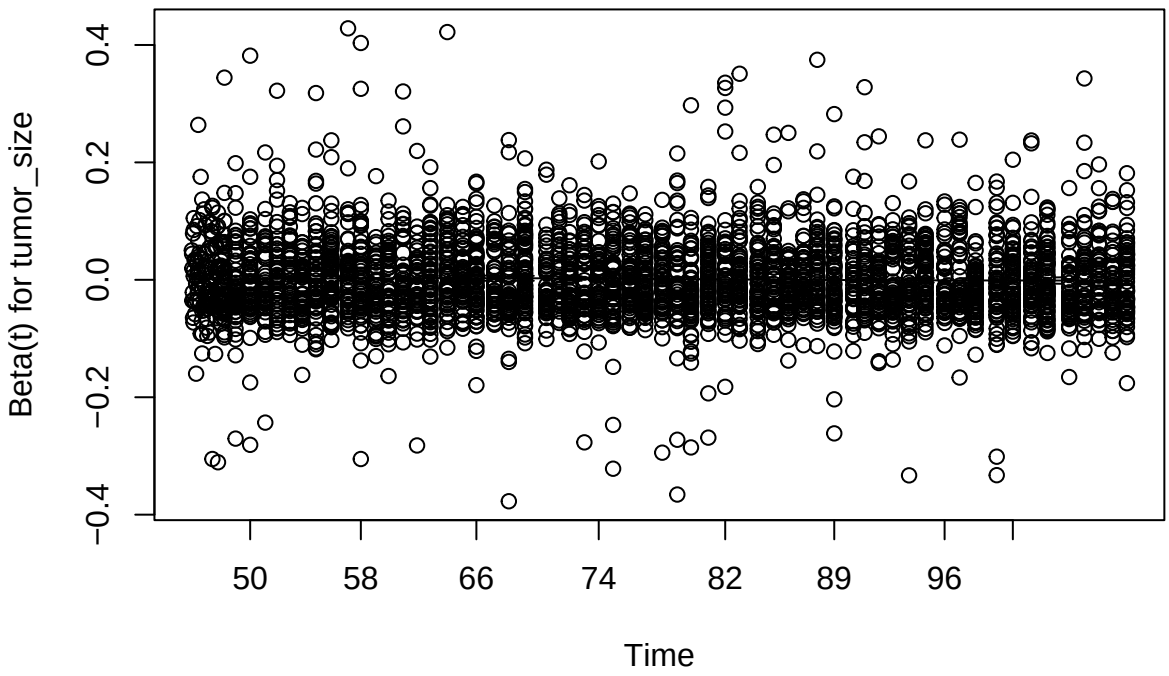

Beta(t) for t_stage
50
0
−50
−150
−250
50
58
66
74
82
89
96
Time

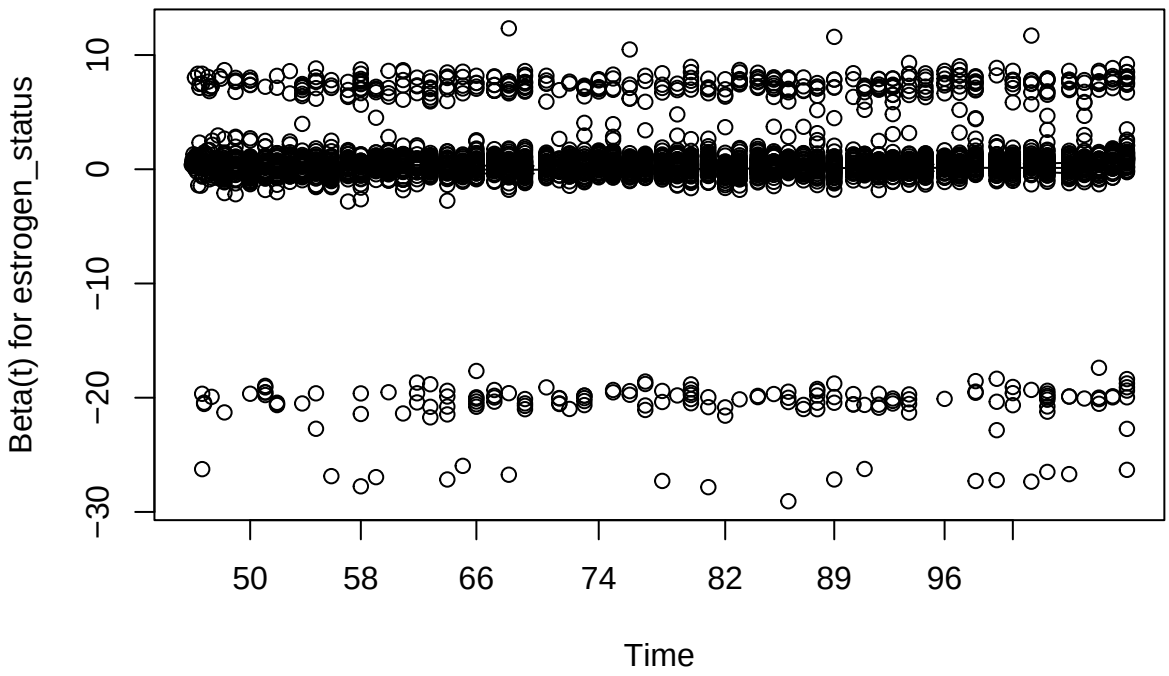

Beta(t) for n_stage
100
50
0
−50
−150
50
58
66
74
82
89
96
Time

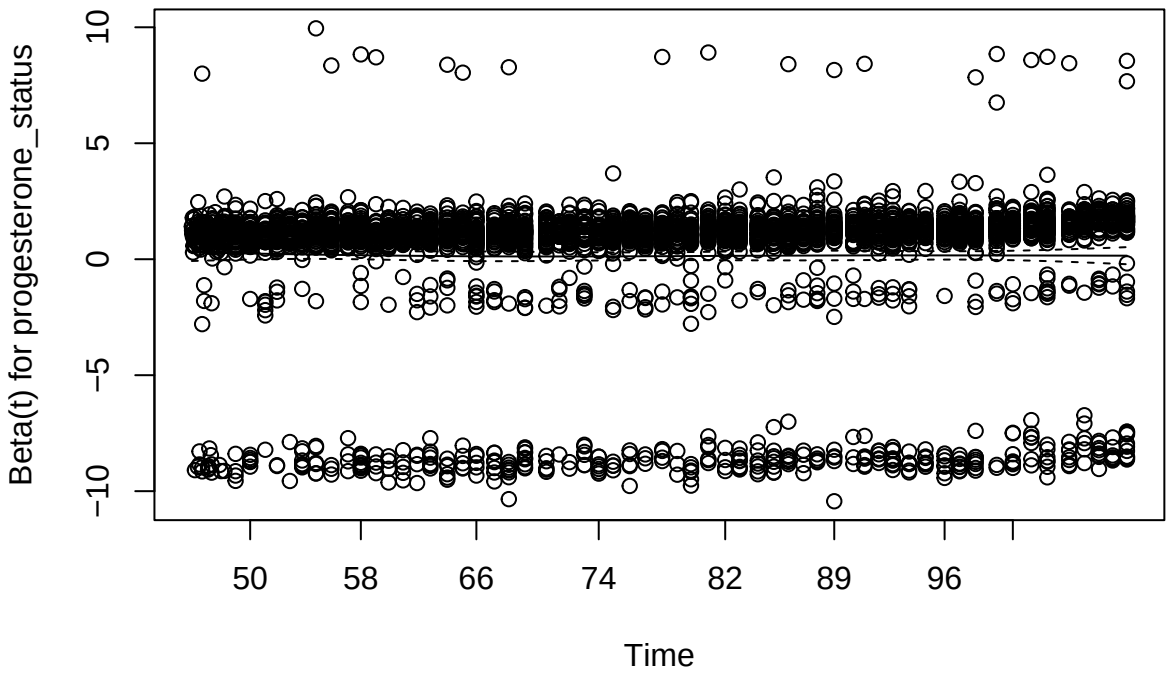

Beta(t) for x6th_stage
100
50
0
−50
−150
50
58
66
74
82
89
96
Time

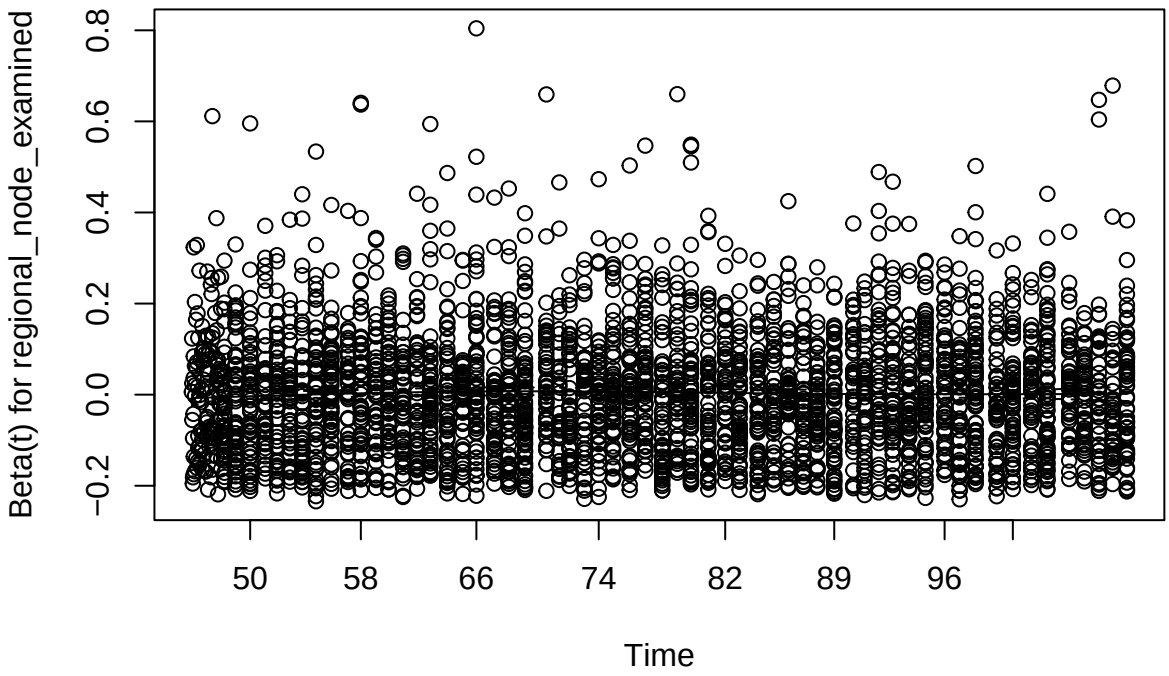

Beta(t) for differentiate
50
0
−50
−150
−250
50
58
66
74
82
89
96
Time

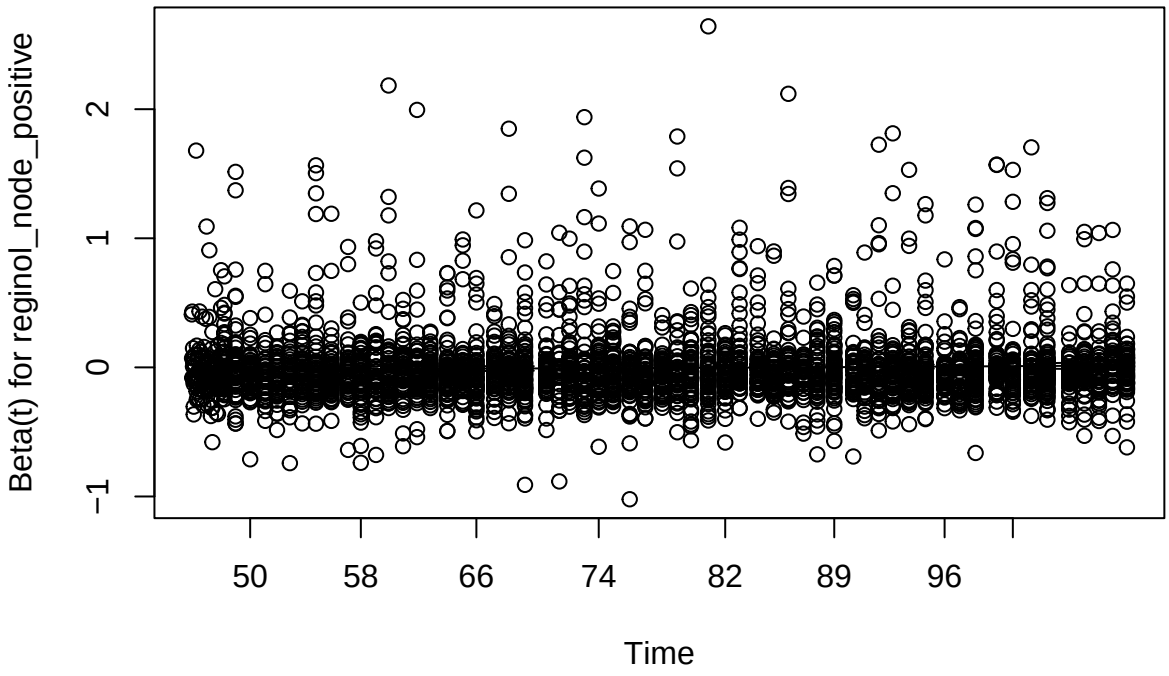

Beta(t) for a_stage
20
0
−20
−40
−60
50
58
66
74
82
89
96
Time


Beta(t) for tumor_size
0.4
0.2
0.0
−0.2
−0.4
50
58
66
74
82
89
96
Time


Beta(t) for estrogen_status
10
0
−10
−20
−30
50
58
66
74
82
89
96
Time


Beta(t) for progesterone_status
10
5
0
−5
−10
50
58
66
74
82
89
96
Time


Beta(t) for regional_node_examined
−0.2
0.0
0.2
0.4
0.6
0.8
50
58
66
74
82
89
96
Time



A non-random pattern or a significant global test (p-value) can indicate violations of the assumption.

Variable Selection

```
stepwise_model <- stepAIC(cox_model, direction = "both", trace = FALSE)
stepwise_model
```

```
## Call:
## coxph(formula = surv_obj ~ differentiate + progesterone_status,
##     data = bc)
##
##                                        coef exp(coef) se(coef)      z
## differentiateModerately differentiated  0.05779   1.05950  0.04116  1.404
## differentiateWell differentiated        0.12600   1.13428  0.05668  2.223
## differentiateUndifferentiated          -0.34872   0.70559  0.31818 -1.096
## progesterone_statusPositive             0.18129   1.19876  0.04921  3.684
##                                            p
## differentiateModerately differentiated 0.16026
## differentiateWell differentiated       0.02622
## differentiateUndifferentiated          0.27307
## progesterone_statusPositive            0.00023
##
## Likelihood ratio test=23.63  on 4 df, p=9.476e-05
## n= 4024, number of events= 3408
```

```r
summary(stepwise_model)
```

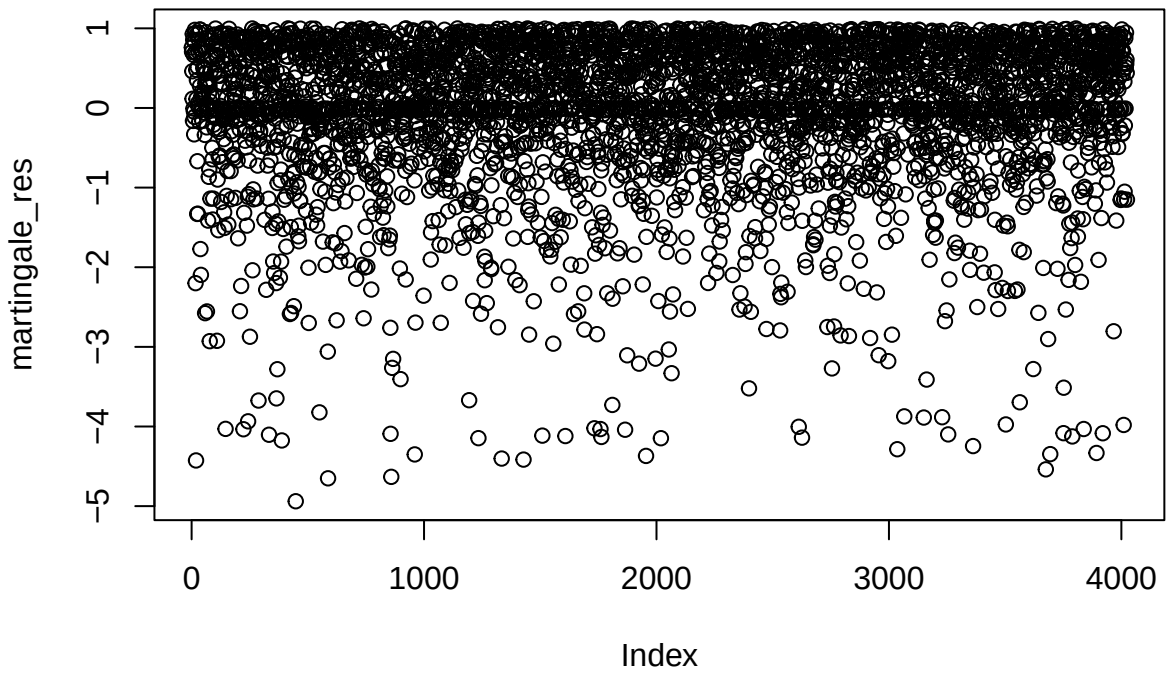
```
## Call:
## coxph(formula = surv_obj ~ differentiate + progesterone_status,
##     data = bc)
##
##   n= 4024, number of events= 3408
##
##                                        coef exp(coef) se(coef)      z
## differentiateModerately differentiated  0.05779   1.05950  0.04116  1.404
## differentiateWell differentiated        0.12600   1.13428  0.05668  2.223
## differentiateUndifferentiated          -0.34872   0.70559  0.31818 -1.096
## progesterone_statusPositive             0.18129   1.19876  0.04921  3.684
##                                       Pr(>|z|)
## differentiateModerately differentiated  0.16026
## differentiateWell differentiated        0.02622 *
## differentiateUndifferentiated           0.27307
## progesterone_statusPositive             0.00023 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                       exp(coef) exp(-coef) lower .95 upper .95
## differentiateModerately differentiated    1.0595     0.9438    0.9774     1.149
## differentiateWell differentiated          1.1343     0.8816    1.0150     1.268
## differentiateUndifferentiated             0.7056     1.4173    0.3782     1.316
## progesterone_statusPositive               1.1988     0.8342    1.0885     1.320
##
## Concordance= 0.525  (se = 0.005 )
## Likelihood ratio test= 23.63  on 4 df,   p=9e-05
## Wald test            = 22.68  on 4 df,   p=1e-04
## Score (logrank) test = 22.76  on 4 df,   p=1e-04
```

After stepwise, it kept two variables differentiate and progesterone status.

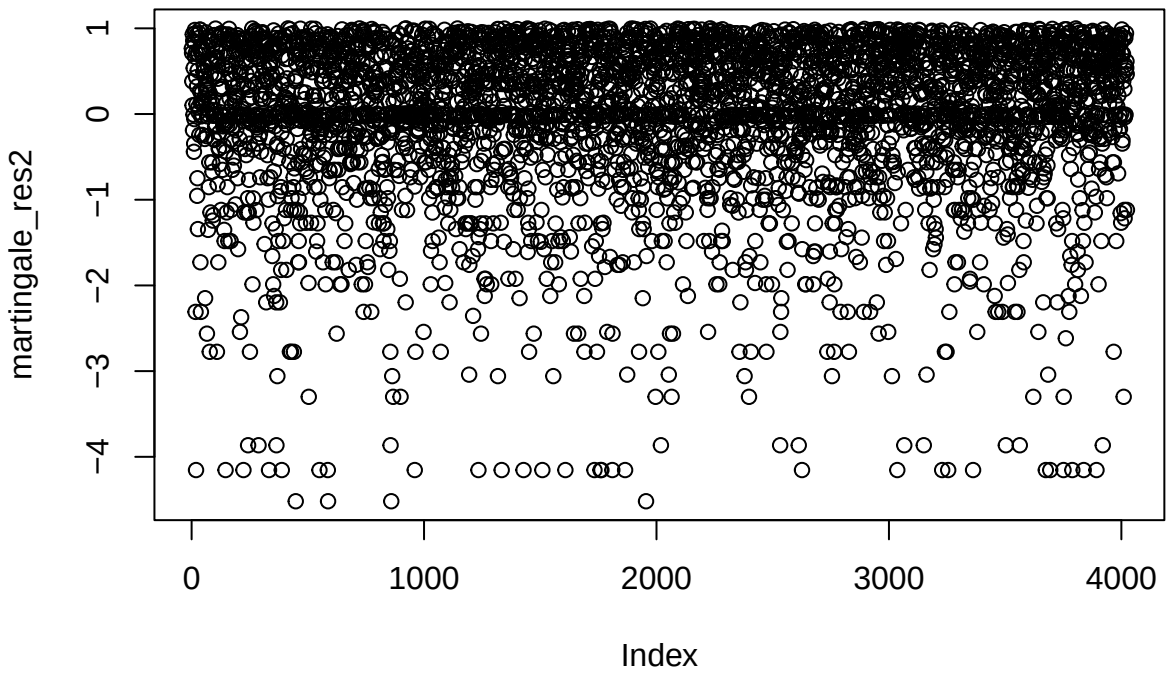
Concordance= 0.525 (se = 0.005)

Model Diagnostics

```r
# Identify Influential Observations
martingale_res <- residuals(cox_model, type = "martingale")
plot(martingale_res)
```




```
martingale_res2 <- residuals(stepwise_model, type = "martingale")
plot(martingale_res2)
```



It seems both of plots have few observations with large negative Martingale residuals, which could be potential outliers or influential observations.

Model Validation

```
y <- bc[["status"]] # Target column
trainIndex <- caret::createDataPartition(y, p = 0.9, list = FALSE)

# Create training and testing sets
train=bc[trainIndex,]
test=bc[-trainIndex,]

# train
surv_obj <- Surv(time = train$survival_months, event = train$status)
fit <- coxph(surv_obj ~ differentiate + progesterone_status, data = train)
summary(fit)
```

```
## Call:
## coxph(formula = surv_obj ~ differentiate + progesterone_status,
##     data = train)
##
##   n= 3622, number of events= 3057
##
##                                       coef exp(coef) se(coef)      z
```

```
## differentiateModerately differentiated  0.05670   1.05834  0.04339  1.307
## differentiateWell differentiated        0.14429   1.15522  0.05949  2.425
## differentiateUndifferentiated          -0.41679   0.65916  0.41005 -1.016
## progesterone_statusPositive             0.20322   1.22534  0.05205  3.904
##                                        Pr(>|z|)
## differentiateModerately differentiated   0.1912
## differentiateWell differentiated         0.0153 *
## differentiateUndifferentiated            0.3094
## progesterone_statusPositive            9.44e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                        exp(coef) exp(-coef) lower .95 upper .95
## differentiateModerately differentiated    1.0583     0.9449    0.9721     1.152
## differentiateWell differentiated          1.1552     0.8656    1.0281     1.298
## differentiateUndifferentiated             0.6592     1.5171    0.2951     1.472
## progesterone_statusPositive               1.2253     0.8161    1.1065     1.357
##
## Concordance= 0.527  (se = 0.006 )
## Likelihood ratio test= 25.93  on 4 df,   p=3e-05
## Wald test            = 24.9  on 4 df,   p=5e-05
## Score (logrank) test = 25.01  on 4 df,   p=5e-05
```

```
# Test data
predictions <- predict(stepwise_model, newdata = test, type = "risk")
perf_metric <- survival::survConcordance(Surv(survival_months, status) ~ predictions,
                                         data = test)$concordance
```

```
## Warning: 'survival::survConcordance' is deprecated.
## Use 'concordance' instead.
## See help("Deprecated")

## Warning: 'survConcordance.fit' is deprecated.
## Use 'concordancefit' instead.
## See help("Deprecated")
```

```
perf_metric
```

```
## concordant
##  0.5082013
```

concordant 0.5336526

```
y <- bc[["status"]] # Target column
trainIndex <- caret::createDataPartition(y, p = 0.9, list = FALSE)

# Create training and testing sets
train=bc[trainIndex,]
test=bc[-trainIndex,]

# train
surv_obj <- Surv(time = train$survival_months, event = train$status)
```

```r
fit <- coxph(surv_obj ~ age + race + marital_status + t_stage + n_stage + x6th_stage +
                    differentiate + grade + a_stage + tumor_size + estrogen_status +
                    progesterone_status + regional_node_examined + reginol_node_positive,
             data = train)
summary(fit)
```

```
## Call:
## coxph(formula = surv_obj ~ age + race + marital_status + t_stage +
##     n_stage + x6th_stage + differentiate + grade + a_stage +
##     tumor_size + estrogen_status + progesterone_status + regional_node_examined +
##     reginol_node_positive, data = train)
##
##   n= 3622, number of events= 3069
##
##                                          coef exp(coef)  se(coef)      z
## age                                  -0.001626  0.998376  0.002169 -0.749
## raceBlack                             0.061370  1.063292  0.075477  0.813
## raceOther                            -0.028633  0.971773  0.065244 -0.439
## marital_statusDivorced               -0.056236  0.945316  0.058352 -0.964
## marital_statusSingle                 -0.069007  0.933320  0.052997 -1.302
## marital_statusWidowed                -0.030296  0.970158  0.081388 -0.372
## marital_statusSeparated               0.209168  1.232652  0.191503  1.092
## t_stageT2                             0.010247  1.010300  0.090014  0.114
## t_stageT3                            -0.092767  0.911406  0.150973 -0.614
## t_stageT4                            -0.317765  0.727774  0.326552 -0.973
## n_stageN2                            -0.149586  0.861065  0.106521 -1.404
## n_stageN3                            -0.077936  0.925024  0.149577 -0.521
## x6th_stageIIIA                        0.125559  1.133782  0.124443  1.009
## x6th_stageIIIC                              NA        NA  0.000000     NA
## x6th_stageIIB                        -0.031439  0.969050  0.095035 -0.331
## x6th_stageIIIB                        0.171439  1.187011  0.346299  0.495
## differentiateModerately differentiated  0.026723  1.027083  0.044362  0.602
## differentiateWell differentiated      0.100247  1.105444  0.061160  1.639
## differentiateUndifferentiated        -0.522040  0.593309  0.337257 -1.548
## grade2                                      NA        NA  0.000000     NA
## grade3                                      NA        NA  0.000000     NA
## gradeanaplastic; Grade IV                   NA        NA  0.000000     NA
## a_stageRegional                      -0.005864  0.994153  0.154764 -0.038
## tumor_size                            0.001130  1.001130  0.001951  0.579
## estrogen_statusPositive               0.133806  1.143171  0.095808  1.397
## progesterone_statusPositive           0.142743  1.153434  0.056897  2.509
## regional_node_examined                0.002473  1.002476  0.002433  1.016
## reginol_node_positive                -0.001219  0.998782  0.008492 -0.143
##                                      Pr(>|z|)
## age                                    0.4536
## raceBlack                              0.4162
## raceOther                              0.6608
## marital_statusDivorced                 0.3352
## marital_statusSingle                   0.1929
## marital_statusWidowed                  0.7097
## marital_statusSeparated                0.2747
## t_stageT2                              0.9094
## t_stageT3                              0.5389
```

```
## t_stageT4                                  0.3305
## n_stageN2                                  0.1602
## n_stageN3                                  0.6023
## x6th_stageIIIA                             0.3130
## x6th_stageIIIC                                 NA
## x6th_stageIIB                              0.7408
## x6th_stageIIIB                             0.6206
## differentiateModerately differentiated     0.5469
## differentiateWell differentiated           0.1012
## differentiateUndifferentiated              0.1216
## grade2                                         NA
## grade3                                         NA
## gradeanaplastic; Grade IV                      NA
## a_stageRegional                            0.9698
## tumor_size                                 0.5627
## estrogen_statusPositive                    0.1625
## progesterone_statusPositive                0.0121 *
## regional_node_examined                     0.3095
## reginol_node_positive                      0.8859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                        exp(coef) exp(-coef) lower .95 upper .95
## age                                       0.9984     1.0016    0.9941     1.003
## raceBlack                                 1.0633     0.9405    0.9171     1.233
## raceOther                                 0.9718     1.0290    0.8551     1.104
## marital_statusDivorced                    0.9453     1.0578    0.8432     1.060
## marital_statusSingle                      0.9333     1.0714    0.8412     1.035
## marital_statusWidowed                     0.9702     1.0308    0.8271     1.138
## marital_statusSeparated                   1.2327     0.8113    0.8469     1.794
## t_stageT2                                 1.0103     0.9898    0.8469     1.205
## t_stageT3                                 0.9114     1.0972    0.6780     1.225
## t_stageT4                                 0.7278     1.3741    0.3837     1.380
## n_stageN2                                 0.8611     1.1614    0.6988     1.061
## n_stageN3                                 0.9250     1.0811    0.6900     1.240
## x6th_stageIIIA                            1.1338     0.8820    0.8884     1.447
## x6th_stageIIIC                                NA         NA        NA        NA
## x6th_stageIIB                             0.9691     1.0319    0.8044     1.167
## x6th_stageIIIB                            1.1870     0.8425    0.6021     2.340
## differentiateModerately differentiated    1.0271     0.9736    0.9416     1.120
## differentiateWell differentiated          1.1054     0.9046    0.9806     1.246
## differentiateUndifferentiated             0.5933     1.6855    0.3063     1.149
## grade2                                        NA         NA        NA        NA
## grade3                                        NA         NA        NA        NA
## gradeanaplastic; Grade IV                     NA         NA        NA        NA
## a_stageRegional                           0.9942     1.0059    0.7340     1.346
## tumor_size                                1.0011     0.9989    0.9973     1.005
## estrogen_statusPositive                   1.1432     0.8748    0.9475     1.379
## progesterone_statusPositive               1.1534     0.8670    1.0317     1.290
## regional_node_examined                    1.0025     0.9975    0.9977     1.007
## reginol_node_positive                     0.9988     1.0012    0.9823     1.016
##
## Concordance= 0.537  (se = 0.006 )
## Likelihood ratio test= 34.41  on 24 df,   p=0.08
```

```
## Wald test            = 32.99  on 24 df,   p=0.1
## Score (logrank) test = 33.13  on 24 df,   p=0.1
```

```
# Test data
predictions <- predict(stepwise_model, newdata = test, type = "risk")
perf_metric <- survival::survConcordance(Surv(survival_months, status) ~ predictions,
                                         data = test)$concordance
```

```
## Warning: 'survival::survConcordance' is deprecated.
## Use 'concordance' instead.
## See help("Deprecated")
```

```
## Warning: 'survConcordance.fit' is deprecated.
## Use 'concordancefit' instead.
## See help("Deprecated")
```

```
perf_metric
```

```
## concordant
##  0.5372427
```

concordant 0.5430644