Multi-Task Neural Learning for Survival Analysis: from traditional to machine learning models

Diego Vallarino

Junio 2022

Contents

---

1 Descritive

```
head(veteran)
```

```
##   trt celltype time status karno diagtime age prior
## 1   1 squamous   72      1    60        7  69     0
## 2   1 squamous  411      1    70        5  64    10
## 3   1 squamous  228      1    60        3  38     0
## 4   1 squamous  126      1    60        9  63    10
## 5   1 squamous  118      1    70       11  65    10
## 6   1 squamous   10      1    20        5  49     0
```

```r
str(veteran)
```

```
## 'data.frame':    137 obs. of  8 variables:
##  $ trt     : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ celltype: Factor w/ 4 levels "squamous","smallcell",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ time    : num  72 411 228 126 118 10 82 110 314 100 ...
##  $ status  : num  1 1 1 1 1 1 1 1 1 0 ...
##  $ karno   : num  60 70 60 60 70 20 40 80 50 70 ...
##  $ diagtime: num  7 5 3 9 11 5 10 29 18 6 ...
##  $ age     : num  69 64 38 63 65 49 69 68 43 70 ...
##  $ prior   : num  0 10 0 10 10 0 10 0 0 0 ...
```

```r
describeBy(veteran, veteran$celltype)
```

```
##
##  Descriptive statistics by group
## group: squamous
##           vars  n   mean     sd median trimmed    mad min max range  skew
## trt          1 35   1.57   0.50      2    1.59   0.00   1   2     1 -0.28
## celltype*    2 35   1.00   0.00      1    1.00   0.00   1   1     0   NaN
## time         3 35 200.20 248.23    111  152.41 142.33   1 999   998  1.91
## status       4 35   0.89   0.32      1    0.97   0.00   0   1     1 -2.32
## karno        5 35  60.86  20.49     60   62.07  14.83  20  90    70 -0.49
## diagtime     6 35  11.03  11.53      7    8.90   5.93   1  58    57  2.32
## age          7 35  58.46  10.37     62   59.03  10.38  35  81    46 -0.40
## prior        8 35   4.00   4.97      0    3.79   0.00   0  10    10  0.39
##           kurtosis    se
## trt          -1.98  0.08
## celltype*      NaN  0.00
## time          3.35 41.96
## status        3.49  0.05
## karno        -0.69  3.46
## diagtime      5.92  1.95
## age          -0.41  1.75
## prior        -1.90  0.84
## ------------------------------------------------------------
## group: smallcell
##           vars  n  mean    sd median trimmed   mad min max range  skew kurtosis
## trt          1 48  1.38  0.49    1.0    1.35  0.00   1   2     1  0.50    -1.79
## celltype*    2 48  2.00  0.00    2.0    2.00  0.00   2   2     0   NaN      NaN
## time         3 48 71.67 85.77   51.0   55.10 50.41   2 392   390  2.35     5.68
## status       4 48  0.94  0.24    1.0    1.00  0.00   0   1     1 -3.50    10.49
## karno        5 48 53.54 19.10   60.0   53.88 29.65  20  85    65 -0.15    -1.25
## diagtime     6 48  9.25 13.91    4.0    6.58  2.97   1  87    86  3.85    17.80
## age          7 48 59.88  9.92   62.5   60.83  8.90  35  72    37 -0.88    -0.25
## prior        8 48  2.29  4.25    0.0    1.75  0.00   0  10    10  1.25    -0.45
##              se
## trt        0.07
## celltype*  0.00
## time      12.38
## status     0.04
## karno      2.76
## diagtime   2.01
```

```
## age        1.43
## prior      0.61
## ------------------------------------------------------------
## group: adeno
##           vars  n  mean    sd median trimmed   mad min max range  skew kurtosis
## trt          1 27  1.67  0.48      2    1.70  0.00   1   2     1 -0.67    -1.61
## celltype*    2 27  3.00  0.00      3    3.00  0.00   3   3     0   NaN      NaN
## time         3 27 64.11 50.59     51   59.70 48.93   3 186   183  0.71    -0.50
## status       4 27  0.96  0.19      1    1.00  0.00   0   1     1 -4.63    20.22
## karno        5 27 58.11 22.12     60   58.70 29.65  10  99    89 -0.20    -0.83
## diagtime     6 27  5.63  4.76      4    4.65  1.48   2  22    20  2.29     4.57
## age          7 27 57.41 11.32     61   57.70  5.93  34  81    47 -0.43    -0.54
## prior        8 27  1.85  3.96      0    1.30  0.00   0  10    10  1.53     0.36
##             se
## trt       0.09
## celltype* 0.00
## time      9.74
## status    0.04
## karno     4.26
## diagtime  0.92
## age       2.18
## prior     0.76
## ------------------------------------------------------------
## group: large
##           vars  n   mean     sd median trimmed    mad min max range  skew
## trt          1 27   1.44   0.51      1    1.43   0.00   1   2     1  0.21
## celltype*    2 27   4.00   0.00      4    4.00   0.00   4   4     0   NaN
## time         3 27 166.11 124.22    156  153.35 111.19  12 553   541  1.13
## status       4 27   0.96   0.19      1    1.00   0.00   0   1     1 -4.63
## karno        5 27  65.00  17.49     70   65.87  14.83  30  90    60 -0.66
## diagtime     6 27   8.15   4.99      8    7.96   5.93   1  18    17  0.27
## age          7 27  56.22  11.16     62   56.87   7.41  37  68    31 -0.57
## prior        8 27   3.70   4.92      0    3.48   0.00   0  10    10  0.51
##           kurtosis    se
## trt          -2.03  0.10
## celltype*      NaN  0.00
## time          1.36 23.91
## status       20.22  0.04
## karno        -0.57  3.37
## diagtime     -1.32  0.96
## age          -1.38  2.15
## prior        -1.81  0.95
```

```
describeBy(veteran, veteran$trt)
```

```
##
##  Descriptive statistics by group
## group: 1
##           vars  n   mean     sd median trimmed   mad min max range  skew
## trt          1 69   1.00   0.00      1    1.00  0.00   1   1     0   NaN
## celltype*    2 69   2.35   1.05      2    2.32  1.48   1   4     3  0.40
## time         3 69 115.14 112.74     97   97.61 97.85   3 553   550  1.55
## status       4 69   0.93   0.26      1    1.00  0.00   0   1     1 -3.23
## karno        5 69  59.20  18.74     60   60.26 29.65  20  90    70 -0.40
```

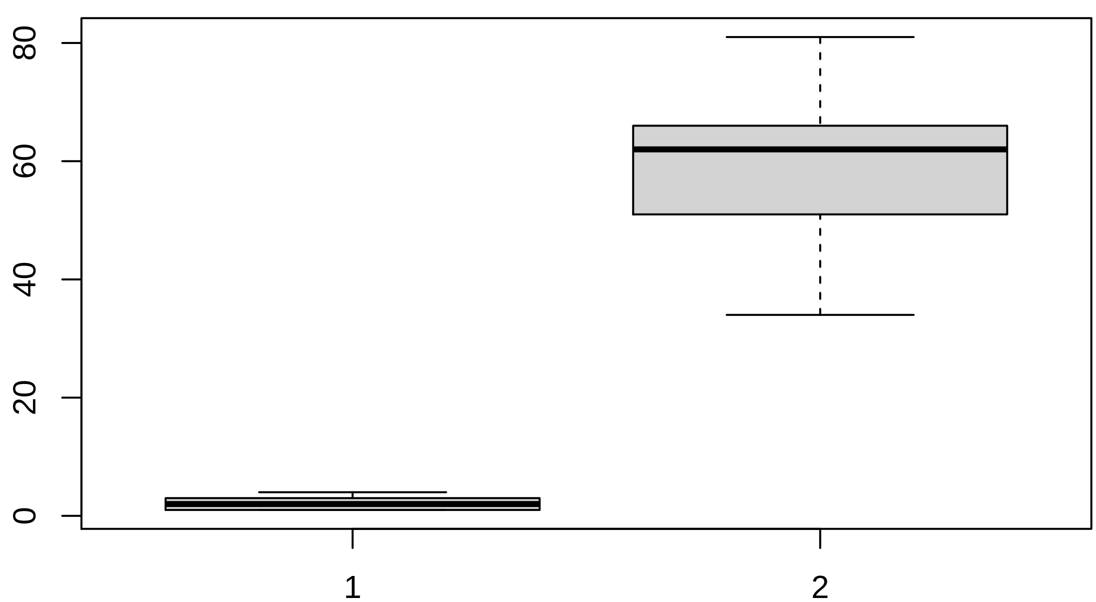
```
## diagtime     6 69   8.65   8.76      5    7.11  4.45   1  58    57  3.02
## age          7 69  57.51  10.81     62   58.21  8.90  34  81    47 -0.54
## prior        8 69   3.04   4.64      0    2.63  0.00   0  10    10  0.83
##           kurtosis    se
## trt            NaN  0.00
## celltype*    -1.09  0.13
## time          2.57 13.57
## status        8.54  0.03
## karno        -0.92  2.26
## diagtime     12.79  1.05
## age          -0.74  1.30
## prior        -1.33  0.56
## ------------------------------------------------------------
## group: 2
##           vars  n   mean     sd median trimmed   mad min max range  skew
## trt          1 68   2.00   0.00    2.0    2.00  0.00   2   2     0   NaN
## celltype*    2 68   2.32   1.09    2.0    2.29  1.48   1   4     3  0.17
## time         3 68 128.21 193.83   52.5   87.16 55.60   1 999   998  2.95
## status       4 68   0.94   0.24    1.0    1.00  0.00   0   1     1 -3.67
## karno        5 68  57.93  21.40   60.0   58.39 29.65  10  99    89 -0.27
## diagtime     6 68   8.90  12.27    4.5    6.52  3.71   1  87    86  4.16
## age          7 68  59.12  10.28   62.0   59.96  8.15  35  81    46 -0.70
## prior        8 68   2.79   4.52    0.0    2.32  0.00   0  10    10  0.96
##           kurtosis    se
## trt            NaN  0.00
## celltype*    -1.30  0.13
## time          9.56 23.50
## status       11.62  0.03
## karno        -0.89  2.59
## diagtime     21.99  1.49
## age          -0.28  1.25
## prior        -1.09  0.55
```

```
describeBy(veteran, veteran$prior)
```

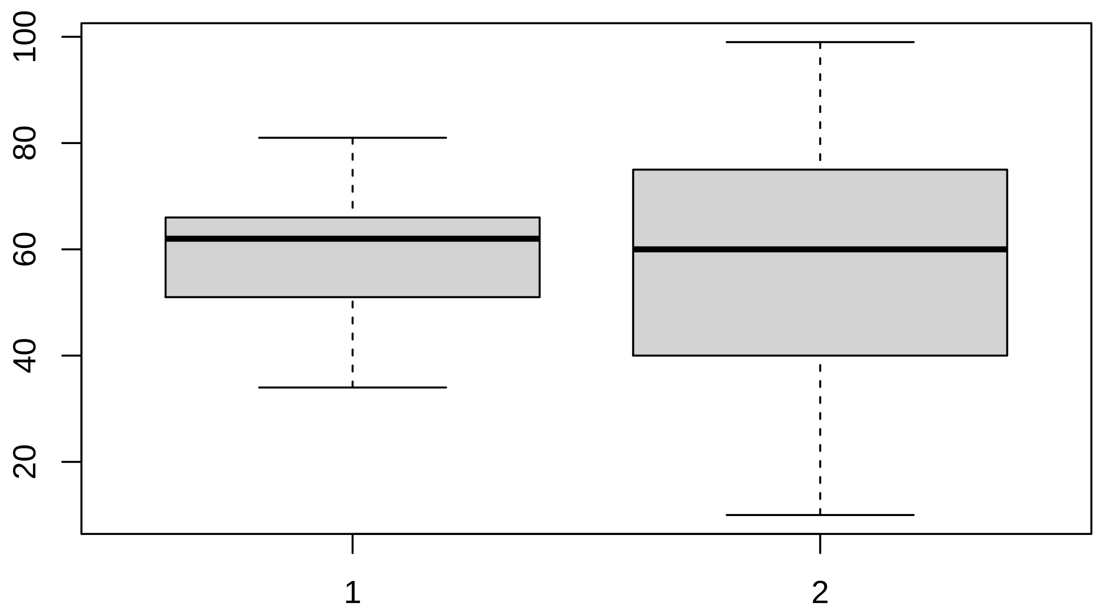
```
##
##  Descriptive statistics by group
## group: 0
##           vars  n   mean     sd median trimmed   mad min max range  skew
## trt          1 97   1.51   0.50      2    1.51  0.00   1   2     1 -0.02
## celltype*    2 97   2.36   1.01      2    2.33  1.48   1   4     3  0.26
## time         3 97 112.15 121.94     80   88.65 78.58   1 587   586  1.88
## status       4 97   0.94   0.24      1    1.00  0.00   0   1     1 -3.58
## karno        5 97  59.32  20.50     60   60.13 29.65  10  99    89 -0.38
## diagtime     6 97   5.94   5.87      4    4.81  2.97   1  36    35  2.77
## age          7 97  58.98  10.71     62   59.82  7.41  34  81    47 -0.70
## prior        8 97   0.00   0.00      0    0.00  0.00   0   0     0   NaN
##           kurtosis    se
## trt          -2.02  0.05
## celltype*    -1.05  0.10
## time          3.42 12.38
## status       10.94  0.02
## karno        -0.77  2.08
## diagtime      9.23  0.60
```

```
## age          -0.32  1.09
## prior          NaN  0.00
## ------------------------------------------------------------
## group: 10
##           vars  n   mean     sd median trimmed   mad min max range  skew
## trt          1 40   1.48   0.51      1    1.47  0.00   1   2     1  0.10
## celltype*    2 40   2.28   1.20      2    2.22  1.48   1   4     3  0.35
## time         3 40 144.60 222.45     69   94.50 84.51   1 999   998  2.78
## status       4 40   0.92   0.27      1    1.00  0.00   0   1     1 -3.11
## karno        5 40  56.75  19.00     60   57.50 29.65  20  90    70 -0.24
## diagtime     6 40  15.65  15.48     12   12.59  6.67   2  87    85  2.94
## age          7 40  56.67  10.07     60   57.47 10.38  36  70    34 -0.51
## prior        8 40  10.00   0.00     10   10.00  0.00  10  10     0   NaN
##           kurtosis    se
## trt          -2.04  0.08
## celltype*    -1.46  0.19
## time          7.80 35.17
## status        7.85  0.04
## karno        -1.13  3.00
## diagtime      9.86  2.45
## age          -1.01  1.59
## prior          NaN  0.00
```

```
boxplot(veteran$celltype, veteran$age)
```



```
boxplot(veteran$age, veteran$karno)
```




```
any(is.na(veteran))
```

```
## [1] FALSE
```

2 Train & Test data

```
set.seed(123)
data.train <- sample_frac(veteran, 0.7)
train_index <- as.numeric(rownames(data.train))
data.test <- veteran [-train_index, ]
```

3 Primeros analisis

```
survdiff(Surv(time, status) ~ trt, data = data.train)
```

```
## Call:
```

```
## survdiff(formula = Surv(time, status) ~ trt, data = data.train)
##
##        N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=1 44       39     33.6     0.851      1.53
## trt=2 52       49     54.4     0.527      1.53
##
##  Chisq= 1.5  on 1 degrees of freedom, p= 0.2
```

```
survdiff(Surv(time, status) ~ celltype, data = data.train)
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ celltype, data = data.train)
##
##                     N Observed Expected (O-E)^2/E (O-E)^2/V
## celltype=squamous  26       23     38.2     6.046     13.02
## celltype=smallcell 36       33     21.0     6.923     10.14
## celltype=adeno     18       17     10.8     3.518      4.22
## celltype=large     16       15     18.0     0.506      0.66
##
##  Chisq= 21.1  on 3 degrees of freedom, p= 1e-04
```

```
survdiff(Surv(time, status) ~ prior + status, data = data.train)
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ prior + status, data = data.train)
##
##                     N Observed Expected (O-E)^2/E (O-E)^2/V
## prior=0, status=0   6        0     5.51   5.50599    6.0703
## prior=0, status=1  63       63    55.94   0.89111    2.5941
## prior=10, status=0  2        0     1.13   1.13051    1.1684
## prior=10, status=1 25       25    25.42   0.00707    0.0109
##
##  Chisq= 7.8  on 3 degrees of freedom, p= 0.05
```

4 Relevancia de las variables a utilizar usando XGBoost

```
require(xgboost)
require(Matrix)
require(data.table)

df <- data.table(data.train, keep.rownames = FALSE)
sparse_matrix <- sparse.model.matrix(status~.-1, data = data.train)
head(sparse_matrix)
```

```
## 6 x 10 sparse Matrix of class "dgCMatrix"
##
## 14   1 1 . . .  25 80  9 52 10
## 50   1 . . 1 . 132 80  5 50  .
## 118 2 . . 1 .  48 10  4 81  .
```

```
## 43   1 . 1 . .  63 50 11 48  .
## 137 2 . . . 1  49 30  3 37  .
## 135 2 . . . 1 231 70 18 67 10
```

```
output_vector = df[,status]

# desarrollo el modelo de relevancia

bst <- xgboost(data = sparse_matrix, label = output_vector, max.depth = 4, eta = 1, nthread = 2, nrounds
```

```
## [10:59:59] WARNING: amalgamation/../src/learner.cc:1115: Starting in XGBoost 1.3.0, the default evalu
## [1]  train-logloss:0.310176
## [2]  train-logloss:0.202168
## [3]  train-logloss:0.148494
## [4]  train-logloss:0.123007
## [5]  train-logloss:0.104148
## [6]  train-logloss:0.089287
## [7]  train-logloss:0.083504
## [8]  train-logloss:0.074024
## [9]  train-logloss:0.067894
## [10] train-logloss:0.064258
```

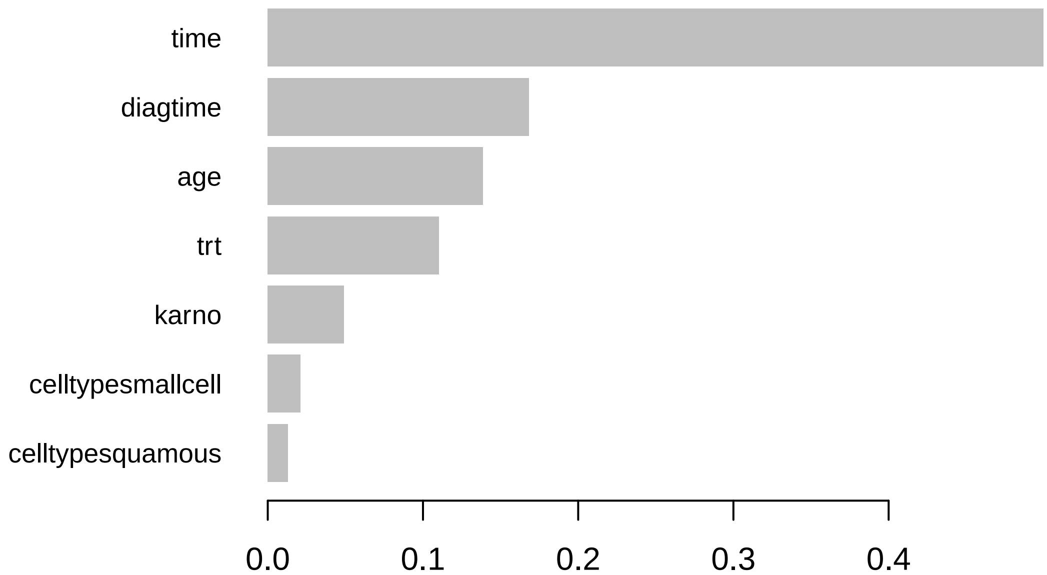
```
# Medimos los la importancia de la variables.

importance <- xgb.importance(feature_names = sparse_matrix@Dimnames[[2]], model = bst)
head(importance)
```

```
##               Feature       Gain      Cover  Frequency
## 1:               time 0.49967836 0.34390974 0.32258065
## 2:           diagtime 0.16815368 0.12704402 0.16129032
## 3:                age 0.13858913 0.14225801 0.19354839
## 4:                trt 0.11037545 0.10794435 0.12903226
## 5:              karno 0.04906092 0.21078392 0.09677419
## 6: celltypesmallcell 0.02099943 0.02759307 0.06451613
```

```
# Mejora en la interpretabilidad de la tabla de datos de importancia de características

importanceRaw <- xgb.importance(feature_names = sparse_matrix@Dimnames[[2]], model = bst, data = sparse_
importanceClean <- importanceRaw[,`:=`(Cover=NULL, Frequency=NULL)]
head(importanceClean)
```

```
##               Feature       Gain
## 1:               time 0.49967836
## 2:           diagtime 0.16815368
## 3:                age 0.13858913
## 4:                trt 0.11037545
## 5:              karno 0.04906092
## 6: celltypesmallcell 0.02099943
```

```
xgb.plot.importance(importance_matrix = importanceRaw)
```



5 Modelo de Tobit - parametric survival model

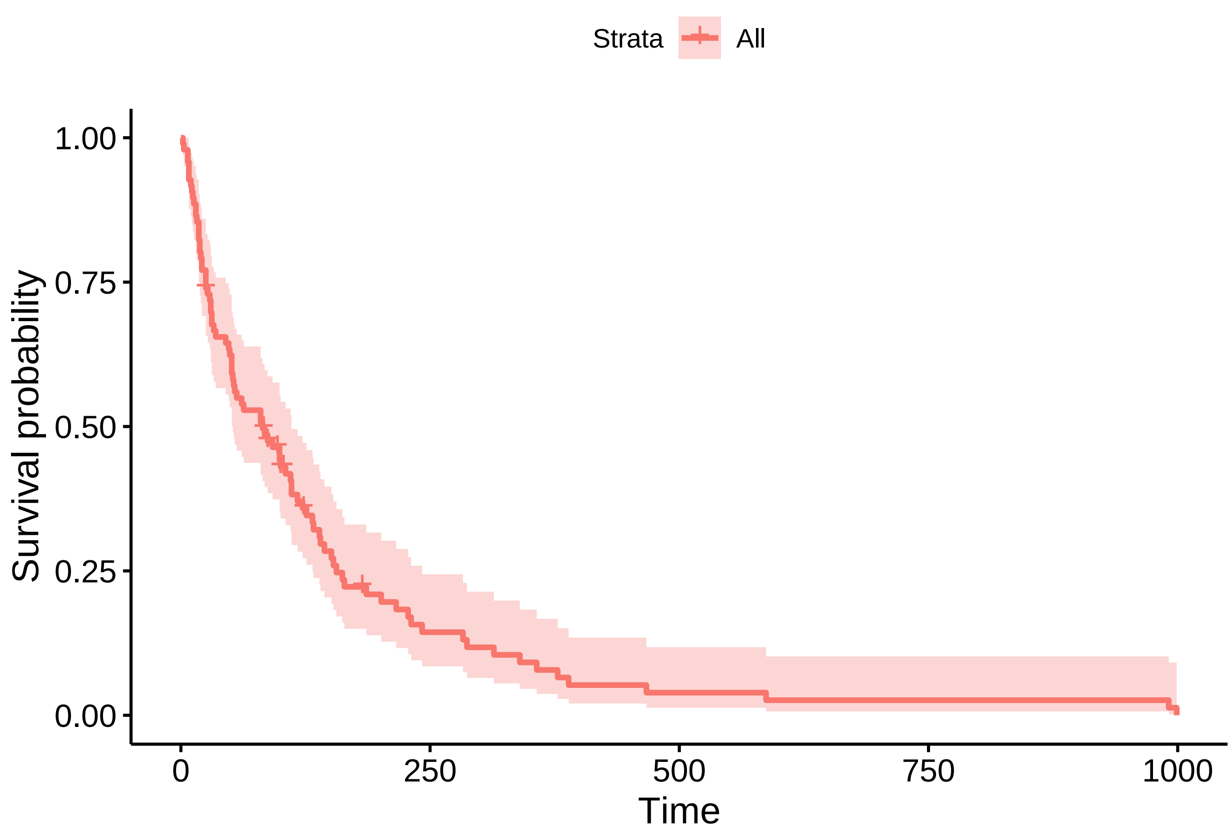
```
surv_obj = Surv(data.test$time, data.test$status)
fit2 <- survreg(Surv(time, status) ~ karno + age + trt, data=data.train)

predictfit2<-predict(fit2, data.test)
metrics_fit2<-Cindex(surv_obj, predicted = predictfit2)

summary(fit2)
```

```
##
## Call:
## survreg(formula = Surv(time, status) ~ karno + age + trt, data = data.train)
##                 Value Std. Error     z       p
## (Intercept)   2.76122    0.75410  3.66 0.00025
## karno         0.03556    0.00519  6.85 7.3e-12
## age          -0.00525    0.01073 -0.49 0.62432
## trt           0.12805    0.20984  0.61 0.54170
## Log(scale)   -0.05965    0.07847 -0.76 0.44718
##
## Scale= 0.942
##
## Weibull distribution
```

```
## Loglik(model)= -496.5   Loglik(intercept only)= -517.3
##  Chisq= 41.4 on 3 degrees of freedom, p= 5.4e-09
## Number of Newton-Raphson Iterations: 5
## n= 96
```

6 Kaplan-Meier Model - non parametric survival model

```
fit3<-survfit(Surv(time, status) ~ 1, data = data.train)
dis_timefit3 = fit3$time

med_indexfit3 = median(1:length(dis_timefit3))

predictfit3<-predictSurvProb(fit3, data.test, dis_timefit3)

metrics_fit3 = Cindex(surv_obj, predicted = predictfit3[, med_indexfit3])

ggsurvplot(fit3, data = veteran, pval = TRUE)
```




```
print(fit3, print.rmean=TRUE)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, data = data.train)
##
```

```
##          n     events     *rmean *se(rmean)     median    0.95LCL    0.95UCL
##         96         88        137         20         82         52        111
##     * restricted mean with upper limit =  999
```

```
summary(fit3, times=c(20, 50, 100, 350))
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, data = data.train)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    20     77      20   0.7917  0.0414       0.7145        0.877
##    50     59      16   0.6234  0.0496       0.5333        0.729
##   100     39      18   0.4300  0.0511       0.3406        0.543
##   350      7      27   0.0916  0.0324       0.0458        0.183
```

7 Cox models - semi parametric survival model

```
fit4 <- coxph(Surv(time, status) ~ ., data=data.train, x = TRUE)
shapiro.test(fit4$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit4$residuals
## W = 0.83037, p-value = 4.037e-09
```

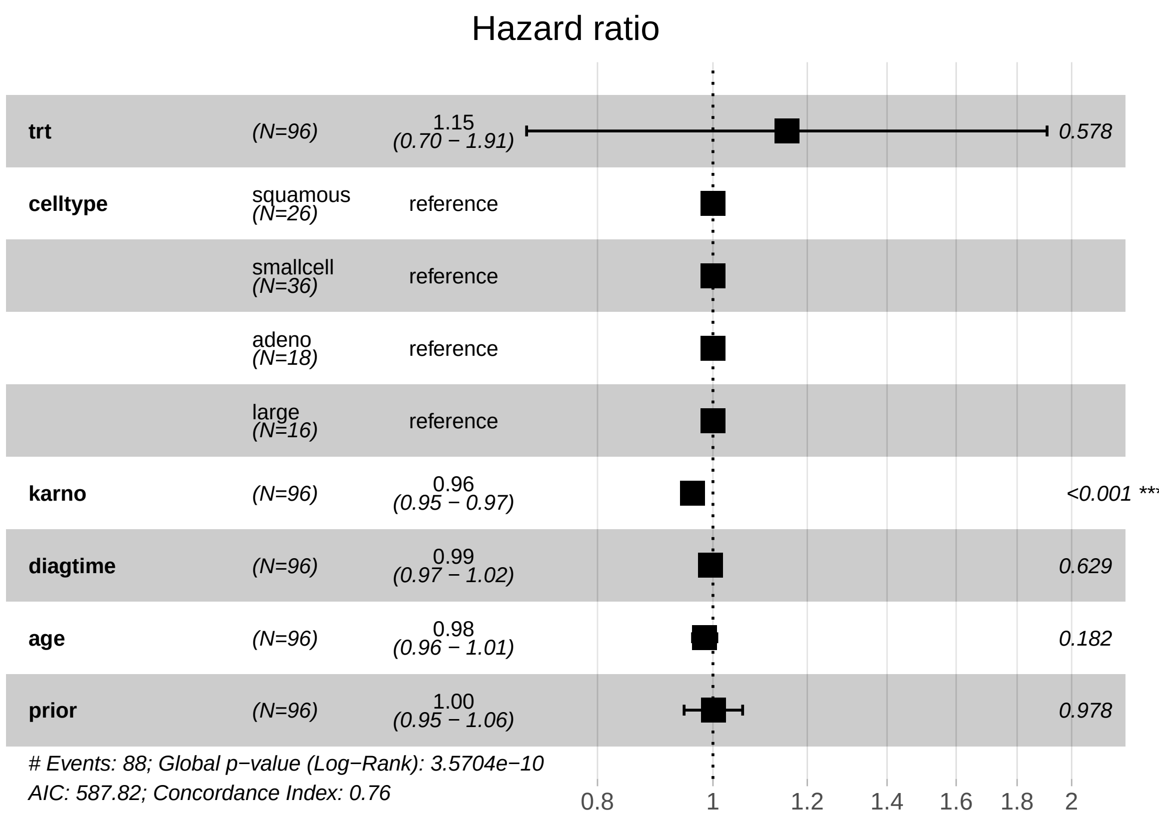
```
anova(fit4)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(time, status)
## Terms added sequentially (first to last)
##
##           loglik   Chisq Df Pr(>|Chi|)
## NULL     -316.20
## trt      -315.47  1.4742  1  0.2246861
## celltype -305.38 20.1670  3  0.0001567 ***
## karno    -286.84 37.0863  1   1.13e-09 ***
## diagtime -286.77  0.1366  1  0.7116438
## age      -285.91  1.7250  1  0.1890525
## prior    -285.91  0.0008  1  0.9779353
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

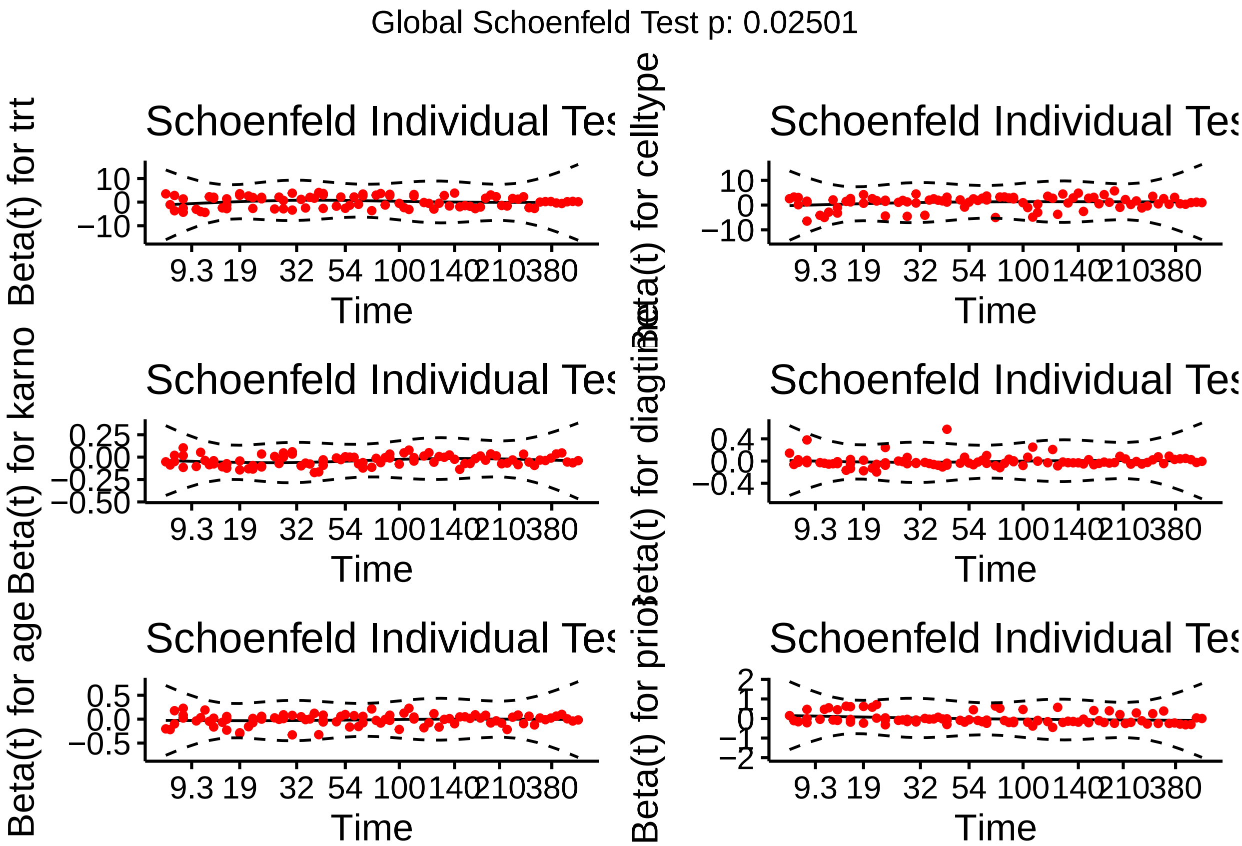
```
dis_timefit4 = fit3$time

med_indexfit4 = median(1:length(dis_timefit4))
predictfit4<-predictSurvProb(fit4, data.test, dis_timefit4)
metrics_fit4 = Cindex(surv_obj, predicted = predictfit4[, med_indexfit4])

summary(fit4)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ ., data = data.train, x = TRUE)
##
##   n= 96, number of events= 88
##
##                 coef  exp(coef)   se(coef)      z Pr(>|z|)
## trt        0.1428218  1.1535242  0.2565892  0.557  0.57779
## celltype1 -0.9161660  0.4000499  0.2341873 -3.912 9.15e-05 ***
## celltype2  0.4729743  1.6047601  0.1992676  2.374  0.01762 *
## celltype3  0.6220558  1.8627536  0.2321140  2.680  0.00736 **
## karno     -0.0394631  0.9613054  0.0064027 -6.164 7.11e-10 ***
## diagtime  -0.0050613  0.9949515  0.0104799 -0.483  0.62913
## age       -0.0159867  0.9841404  0.0119828 -1.334  0.18216
## prior      0.0007983  1.0007986  0.0288504  0.028  0.97793
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## trt          1.1535     0.8669    0.6976    1.9074
## celltype1    0.4000     2.4997    0.2528    0.6331
## celltype2    1.6048     0.6231    1.0859    2.3715
## celltype3    1.8628     0.5368    1.1819    2.9358
## karno        0.9613     1.0403    0.9493    0.9734
## diagtime     0.9950     1.0051    0.9747    1.0156
## age          0.9841     1.0161    0.9613    1.0075
## prior        1.0008     0.9992    0.9458    1.0590
##
## Concordance= 0.759  (se = 0.029 )
## Likelihood ratio test= 60.59  on 8 df,   p=4e-10
## Wald test            = 55.67  on 8 df,   p=3e-09
## Score (logrank) test = 60.99  on 8 df,   p=3e-10
```

```
ggforest(fit4)
```




```
test_cox<-cox.zph(fit4)
ggcoxzph(test_cox)
```



8 MTLR Model - machine learning model

```
fit7 <- mtlr(Surv(time, status)~., data = data.train, nintervals = 9)
dis_timefit7 = fit7$time_points
med_indexfit7 = median(1:length(dis_timefit7))

predictfit7<-predict(fit7, data.test, type = "mean_time")
metrics_fit7 = Cindex(surv_obj, predicted = predictfit7)

fit7
```

```
##
## Call:  mtlr(formula = Surv(time, status) ~ ., data = data.train, nintervals = 9)
##
## Time points:
##  [1]  11.6  19.0  26.8  46.6  56.9  87.0 107.3 133.5 177.1 296.8
##
##
## Weights:
##            Bias      trt  celltype1  celltype2  celltype3   karno  diagtime      age
## 11.64    0.1171 -0.03339    0.03081   0.000274    0.03571 -0.0197   0.03465 -0.01826
## 19      -0.0507 -0.03835    0.00333  -0.007206    0.00663 -0.0780   0.01243 -0.02928
## 26.82   -0.1685  0.01514   -0.00729   0.014363   -0.00533 -0.1320   0.00240 -0.03835
```

```
## 46.64  -0.0351  0.01607   0.01185  0.031709   0.03268 -0.1219 -0.02626 -0.02049
## 56.91   0.4227  0.01776  -0.03017  0.038213   0.02317 -0.1353  0.01017 -0.01955
## 87     -0.1226  0.03394  -0.02180  0.059672   0.05490 -0.1428  0.00457  0.00763
## 107.27 -0.2863  0.02987  -0.05675  0.059685   0.03929 -0.1099 -0.00350  0.00468
## 133.55 -0.0104  0.00147  -0.04636  0.049062   0.02854 -0.0682  0.00628 -0.00473
## 177.09 -0.0166 -0.04860  -0.06836  0.057887   0.03886 -0.0690 -0.00596  0.03760
## 296.82 -0.0524 -0.04662  -0.05469  0.035507   0.02726 -0.0490  0.01090  0.01674
##           prior
## 11.64   0.01958
## 19      0.05089
## 26.82   0.06572
## 46.64   0.01226
## 56.91   0.00542
## 87      0.01663
## 107.27  0.00601
## 133.55 -0.01814
## 177.09 -0.02078
## 296.82 -0.00646
```

9 Cantidad de observaciones censuradas a la derecha

```
table(fit3$n.censor)
```

```
##
##  0  1
## 68  8
```

```
fit2$y
```

```
##   14   50  118   43  137  135   90   91  130   57   92  121    9   93   99   72
##  25+  132   48   63   49  231   25 103+   15  216   21  186  314   13   99  87+
##   26    7   42  125   83   36   78   81  134  103   76   15   32  106  120  132
##   16   82    7   80  467  287  587   25  111   25  111   11  139   51  140  340
##   41   74   23   27   60   53  107  100  102   96   38   89   34   69  122  111
##   54  242  153  151   12    3   29    8   61   20   51   15   31  100   84   31
##   63   13   82   25   95   21   79  105   47  101   16    6  129   39   31  136
##  156  144  357  117    2 123+  389   80   92   99   30   10   53  122   18  378
##  124    4   88  127   86   52   22  109   70  112   35   40   48   30   12   75
##   45  126  283  164   30  162  97+   18  999   51   52   27   35   21    8  991
##  128   46   80   94  133   29   66  123    3   64  110   84   37    8   10  119
##   19    8   33   87  133   56  105   19  228 182+  83+  201   18  110 100+    7
```

```
fit4$y
```

```
##   14   50  118   43  137  135   90   91  130   57   92  121    9   93   99   72
##  25+  132   48   63   49  231   25 103+   15  216   21  186  314   13   99  87+
##   26    7   42  125   83   36   78   81  134  103   76   15   32  106  120  132
##   16   82    7   80  467  287  587   25  111   25  111   11  139   51  140  340
##   41   74   23   27   60   53  107  100  102   96   38   89   34   69  122  111
```

```
##    54   242   153   151    12     3    29     8    61    20    51    15    31   100    84    31
##    63    13    82    25    95    21    79   105    47   101    16     6   129    39    31   136
##   156   144   357   117     2  123+   389    80    92    99    30    10    53   122    18   378
##   124     4    88   127    86    52    22   109    70   112    35    40    48    30    12    75
##    45   126   283   164    30   162   97+    18   999    51    52    27    35    21     8   991
##   128    46    80    94   133    29    66   123     3    64   110    84    37     8    10   119
##    19     8    33    87   133    56   105    19   228  182+   83+   201    18   110  100+     7
```

```
fit7$response
```

```
##    14    50   118    43   137   135    90    91   130    57    92   121     9    93    99    72
##   25+   132    48    63    49   231    25  103+    15   216    21   186   314    13    99   87+
##    26     7    42   125    83    36    78    81   134   103    76    15    32   106   120   132
##    16    82     7    80   467   287   587    25   111    25   111    11   139    51   140   340
##    41    74    23    27    60    53   107   100   102    96    38    89    34    69   122   111
##    54   242   153   151    12     3    29     8    61    20    51    15    31   100    84    31
##    63    13    82    25    95    21    79   105    47   101    16     6   129    39    31   136
##   156   144   357   117     2  123+   389    80    92    99    30    10    53   122    18   378
##   124     4    88   127    86    52    22   109    70   112    35    40    48    30    12    75
##    45   126   283   164    30   162   97+    18   999    51    52    27    35    21     8   991
##   128    46    80    94   133    29    66   123     3    64   110    84    37     8    10   119
##    19     8    33    87   133    56   105    19   228  182+   83+   201    18   110  100+     7
```

10 Conclusion

```
metrics_fit2
```

```
##   C index
## 0.6857319
```

```
metrics_fit3
```

```
## C index
## 0.50246
```

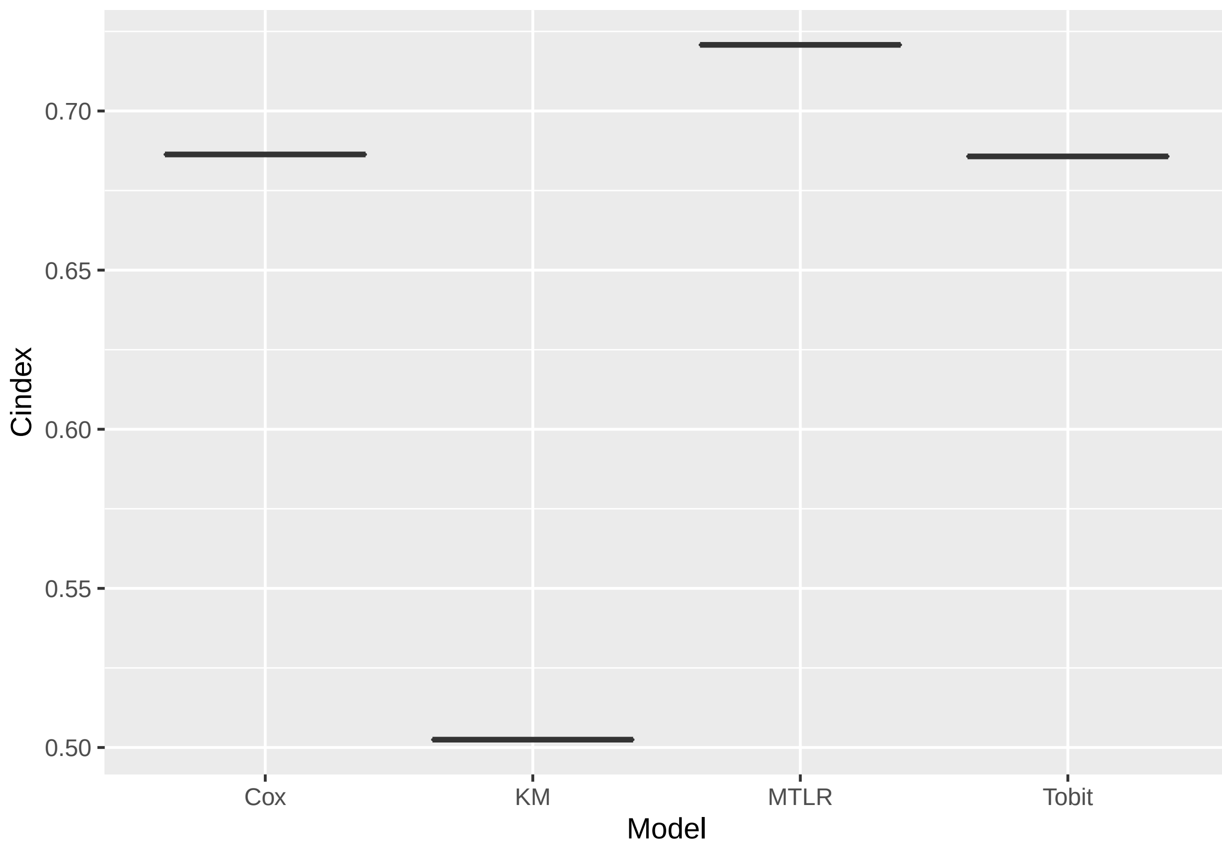
```
metrics_fit4
```

```
##   C index
## 0.6863469
```

```
metrics_fit7
```

```
##   C index
## 0.7207872
```

```
data_CI = data.frame(Cindex = c(metrics_fit2, metrics_fit3, metrics_fit4, metrics_fit7),
                     Model = c(rep('Tobit', 1), rep('KM', 1),rep('Cox', 1),rep('MTLR', 1)))

ggplot(data_CI, aes(x = Model, y = Cindex)) + geom_boxplot()
```



0.70
0.65
0.60
0.55
0.50
Cindex
Cox
KM
MTLR
Tobit
Model