Stochastic Models Assignment 3

16343261

21/07/2023

Before we begin this assignment, I will load in the data as required.

```
library(emplik)
library(survival)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data(myeloma)
colnames(myeloma) <- c("time","vstatus","logBUN","HGB","platelet","age",
                       "logWBC","FRAC","logPBM","protein","SCALC")
myeloma <- as.data.frame(myeloma)
```

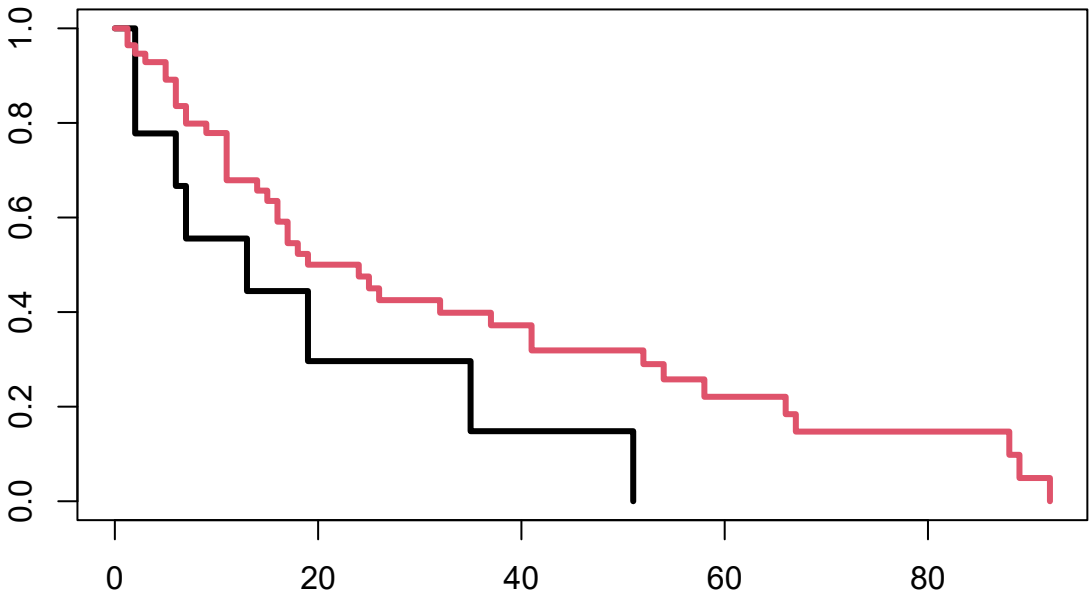
Question 1.

To answer this question, we must first fit our survival object and fit our Kaplan-Meier estimates.

```
survdat <- Surv(time= myeloma$time,event= myeloma$vstatus)
fit <- survfit(survdat~myeloma$platelet, data=myeloma,se=TRUE)
```

Now we will plot our survival curve and interpret.

```
plot(fit,conf.int=FALSE,col=1:2, lwd = 3)
```



Here we see our survival curves with the black line representing patients with abnormal platelets and our red line representing patients with normal platelets. Here we see as time increases, the distances between our curves expands. This expansion implies that there is a clear different time to death for each of our 2 groups. From our curve, it is suggested that those with normal platelets have a considerably longer time to death than those with abnormal platelets.

Question 2.

To compare our median time to death for patients from the two groups, we will use our summary function. Once our survival variable falls below 0.5, the subsequent time value represents our closest estimate of the median time to death.

```
summary(fit)
```

```
## Call: survfit(formula = survdat ~ myeloma$platelet, data = myeloma,
##     se = TRUE)
##
##                 myeloma$platelet=0
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     2      9       2    0.778   0.139       0.5485        1.000
##     6      7       1    0.667   0.157       0.4200        1.000
##     7      6       1    0.556   0.166       0.3097        0.997
##    13      5       1    0.444   0.166       0.2141        0.923
##    19      3       1    0.296   0.164       0.1003        0.875
```

```
##    35      2       1    0.148   0.133       0.0255        0.860
##    51      1       1    0.000     NaN           NA           NA
##
##                 myeloma$platelet=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   1.25     56       2   0.9643  0.0248      0.91689        1.000
##   2.00     54       1   0.9464  0.0301      0.88925        1.000
##   3.00     53       1   0.9286  0.0344      0.86351        0.999
##   5.00     50       2   0.8914  0.0419      0.81301        0.977
##   6.00     48       3   0.8357  0.0501      0.74305        0.940
##   7.00     45       2   0.7986  0.0543      0.69887        0.912
##   9.00     40       1   0.7786  0.0565      0.67534        0.898
##  11.00     39       5   0.6788  0.0645      0.56337        0.818
##  14.00     31       1   0.6569  0.0661      0.53936        0.800
##  15.00     30       1   0.6350  0.0674      0.51573        0.782
##  16.00     29       2   0.5912  0.0695      0.46954        0.744
##  17.00     26       2   0.5457  0.0712      0.42258        0.705
##  18.00     24       1   0.5230  0.0718      0.39964        0.684
##  19.00     23       1   0.5002  0.0722      0.37704        0.664
##  24.00     20       1   0.4752  0.0728      0.35203        0.642
##  25.00     19       1   0.4502  0.0731      0.32749        0.619
##  26.00     18       1   0.4252  0.0732      0.30344        0.596
##  32.00     16       1   0.3986  0.0733      0.27802        0.572
##  37.00     15       1   0.3721  0.0731      0.25319        0.547
##  41.00     14       2   0.3189  0.0716      0.20532        0.495
##  52.00     11       1   0.2899  0.0708      0.17969        0.468
##  54.00      9       1   0.2577  0.0698      0.15151        0.438
##  58.00      7       1   0.2209  0.0689      0.11987        0.407
##  66.00      6       1   0.1841  0.0665      0.09065        0.374
##  67.00      5       1   0.1473  0.0626      0.06403        0.339
##  88.00      3       1   0.0982  0.0579      0.03093        0.312
##  89.00      2       1   0.0491  0.0452      0.00808        0.298
##  92.00      1       1   0.0000     NaN           NA           NA
```

From this summary, we find that the median time to death for a multiple myeloma patient with normal platelets is 24 years. This contrasts sharply with those who has abnormal platelets as their median time to death is 13 years. It is would appear that those with abnormal platelets die earlier than those with normal platelets. This supports our findings in question 1.

To conduct a log-rank test, we must use the survdiff function. Our hypothesis' are as follows:

- H0: Our curves aren't significantly different.
- H1: Our curves are significantly different.

```
diff<- survdiff(survdat~myeloma$platelet, data=myeloma)
diff
```

```
## Call:
## survdiff(formula = survdat ~ myeloma$platelet, data = myeloma)
##
##                       N Observed Expected (O-E)^2/E (O-E)^2/V
## myeloma$platelet=0  9        8     4.46     2.815      3.29
## myeloma$platelet=1 56       40    43.54     0.288      3.29
##
##  Chisq= 3.3  on 1 degrees of freedom, p= 0.07
```

The log-rank test indicates that there is a borderline significant difference in survival between patients with abnormal platelets and normal platelets, but the result is not statistically significant at the 0.05 level. Therefore, at a 95% confidence level, we cannot reject our null hypothesis that our two curves aren't different. This difference could be the result of noise.

Question 3.

To answer this question, we will first fit our Cox proportional hazards model as requested and interpret our results.

```
#Fit the Cox PH model to the data
fit_cox <- coxph(survdat~logBUN + HGB + platelet + age + logWBC +
FRAC + logPBM + protein + SCALC, data=myeloma)
```

We will now perform stepwise variable selection based on AIC (Akaike information criterion) for our Cox proportional hazards model using the step() function along with the AIC criterion. The k argument specifies the number of parameters penalized by AIC, and the trace argument controls whether the stepwise selection process is displayed.

```
# Perform stepwise variable selection using AIC
final_model <- step(fit_cox, direction = "both",
                    scope = formula(fit_cox),
                    k = 2, trace = FALSE)
```

```
summary(final_model)
```

```
## Call:
## coxph(formula = survdat ~ logBUN + HGB, data = myeloma)
##
##   n= 65, number of events= 48
##
##            coef exp(coef) se(coef)      z Pr(>|z|)
## logBUN  1.71597   5.56209  0.61855  2.774  0.00553 **
## HGB    -0.11966   0.88722  0.05742 -2.084  0.03717 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## logBUN    5.5621     0.1798    1.6547   18.6961
## HGB       0.8872     1.1271    0.7928    0.9929
##
## Concordance= 0.675  (se = 0.043 )
## Likelihood ratio test= 12.27  on 2 df,   p=0.002
## Wald test            = 12.51  on 2 df,   p=0.002
## Score (logrank) test = 13.07  on 2 df,   p=0.001
```

Here we find 2 statistically significant variables in LogBUN (log percentage of plasma cells in bone marrow) and HGB (hemoglobin at diagnosis).

- For logBUN, the estimated coefficient is 1.72, and its hazard ratio is 5.56. The hazard ratio indicates that for every one-unit increase in the natural logarithm of BUN, the hazard (risk

of an event) increases by a factor of approximately 5.56. This is a considerable finding and illustrates that the log percentage of plasma cells in bone marrow has a considerable negative affect on the time to death of a patient.

- For HGB, the estimated coefficient is -0.12, and its hazard ratio is 0.89. The hazard ratio indicates that for every one-unit decrease in hemoglobin (HGB) level, the hazard (risk of an event) decreases by a factor of approximately 0.89. Unlike log percentage of plasma cells in bone marrow, hemoglobin at diagnosis heavily increases one's survival chances.