qbs121_hw7_gibran

Gibran Erlangga

2/28/2022

Data Analyses

2.1 PBC (50 points)

Use the dataset pbc that comes with the library survival in R. The time and status indicators compose the endpoint.

```
library(survival)
head(pbc)
```

```
##   id time status trt      age sex ascites hepato spiders edema bili chol
## 1  1  400      2   1 58.76523   f       1      1       1   1.0 14.5  261
## 2  2 4500      0   1 56.44627   f       0      1       1   0.0  1.1  302
## 3  3 1012      2   1 70.07255   m       0      0       0   0.5  1.4  176
## 4  4 1925      2   1 54.74059   f       0      1       1   0.5  1.8  244
## 5  5 1504      1   2 38.10541   f       0      1       1   0.0  3.4  279
## 6  6 2503      2   2 66.25873   f       0      1       0   0.0  0.8  248
##   albumin copper alk.phos    ast trig platelet protime stage
## 1    2.60    156   1718.0 137.95  172      190    12.2     4
## 2    4.14     54   7394.8 113.52   88      221    10.6     3
## 3    3.48    210    516.0  96.10   55      151    12.0     4
## 4    2.54     64   6121.8  60.63   92      183    10.3     4
## 5    3.53    143    671.0 113.15   72      136    10.9     3
## 6    3.98     50    944.0  93.00   63       NA    11.0     3
```

1. Competing Risks: The variable status has 3 levels, 0 for censoring, 1 for transplant, 2 for death.

(a) (2pts) If someone had a transplant and died sometime later what would this data record?

The status should show 2 for someone who had transplant and died when the study is running.

(b) Download the library, cmprsk

```
library(cmprsk)
```

```
## Warning: package 'cmprsk' was built under R version 4.1.1
```

(c) (3) Run the command to get "cumuative incidence" o=with(pbc, cuminc(time,status))

```
o <- with(pbc, cuminc(time, status))
o
```

```
## Estimates and Variances:
## $est
##              1000       2000      3000       4000
## 1 1 0.01681293 0.04457393 0.0698984 0.08382977
## 1 2 0.18239707 0.30198210 0.4152135 0.56435900
##
## $var
##               1000        2000         3000         4000
## 1 1 3.981527e-05 0.0001134119 0.0002123444 0.0003017893
## 1 2 3.590633e-04 0.0005594984 0.0008438443 0.0015692409
```

(d) (5) Plot the output, plot(o), and interpret the resulting curves

```
plot(o)
```
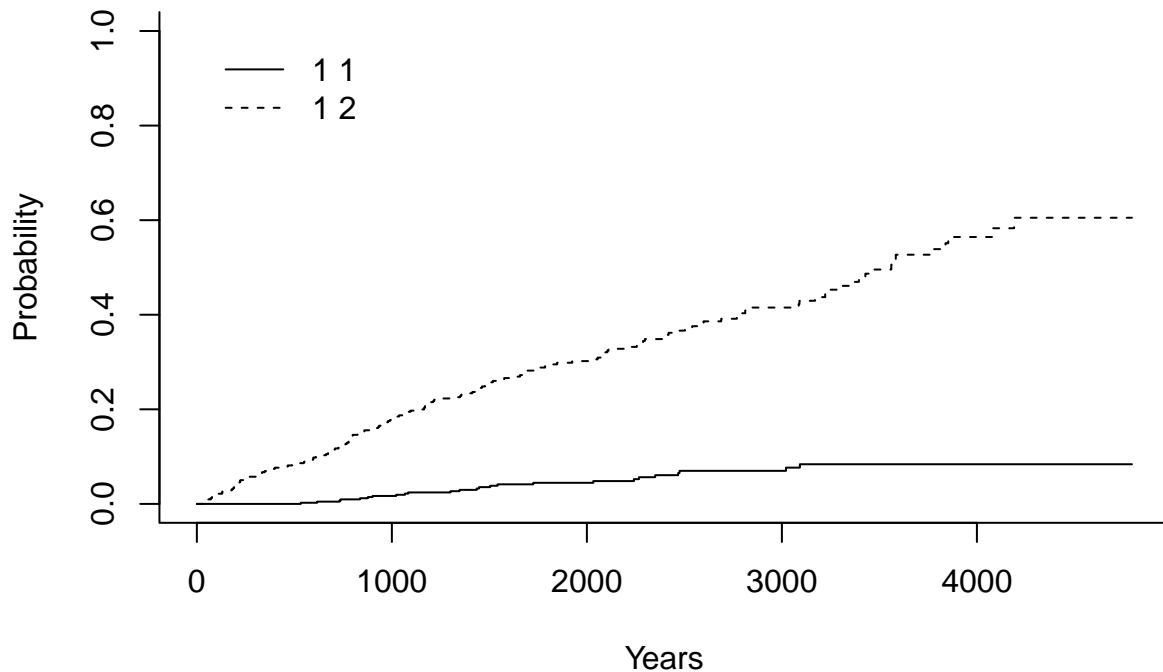

2. (2) What proportion of patients are censored?

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------- tidyverse 1.3.1 --
```

2

```
## v ggplot2 3.3.5          v purrr   0.3.4
## v tibble  3.1.5          v dplyr   1.0.7.9000
## v tidyr   1.1.4          v stringr 1.4.0
## v readr   2.0.2          v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
## Warning: package 'tibble' was built under R version 4.1.1
```

```
## Warning: package 'tidyr' was built under R version 4.1.1
```

```
## Warning: package 'readr' was built under R version 4.1.1
```

```
## Warning: package 'stringr' was built under R version 4.1.1
```

```
## Warning: package 'forcats' was built under R version 4.1.1
```

```
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
pbc %>%
group_by(pbc$status == 0) %>%
summarise(n=n()) %>%
mutate(freq = n/sum(n))
```

```
## # A tibble: 2 x 3
##   `pbc$status == 0`     n  freq
##   <lgl>             <int> <dbl>
## 1 FALSE               186 0.445
## 2 TRUE                232 0.555
```

3. (3) What is the total follow-up time in this study?

```
sum(pbc$time)
```

```
## [1] 801633
```

4. (5) Calculate the incidence rate of transplant and death (e.g. number of events divided by total follow-up time)

```
sum(pbc$status == 1)/sum(pbc$time)
```

```
## [1] 3.118634e-05
```

5. Evaluate if trt arm impacts the composite endpoint, transplant or death by

(a) (2) first coding the composite endpoint, tr.death=status> 0

```
tr.death <- pbc$status > 0
```
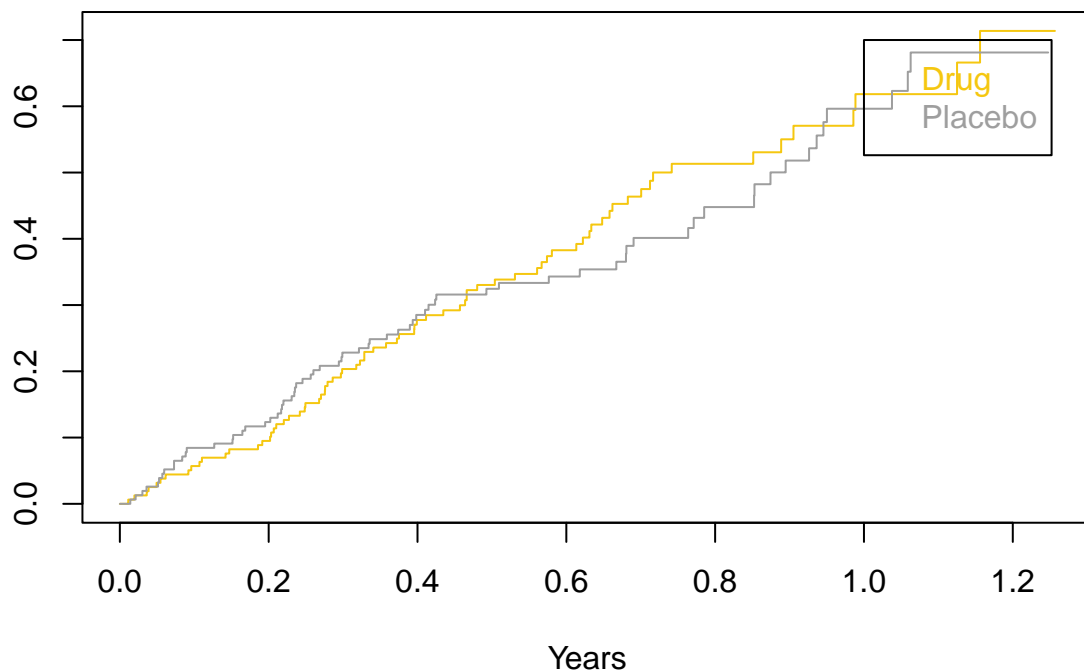
(b) (3) reporting the number of events and rate in each arm

```
tapply(tr.death, pbc$trt, sum)
```

```
##  1  2
## 75 69
```

(c) (5) plotting survival curves and reporting median time to transplant or death in each treatment group

```
plot(survfit(Surv(time/3625.5, tr.death) ~ trt, data = pbc),
     fun = "event",
     col=7:8,
     xlab="Years")

legend(1,0.7,
       text.col=7:8,
       legend=c("Drug","Placebo"))
```


(d) (5) conducting a log-rank test

```
survdiff(Surv(time, tr.death) ~ trt, data = pbc)
```

```
## Call:
## survdiff(formula = Surv(time, tr.death) ~ trt, data = pbc)
##
## n=312, 106 observations deleted due to missingness.
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=1 158       75     72.9    0.0595     0.121
## trt=2 154       69     71.1    0.0611     0.121
##
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.7
```

(e) (5) conducting a Peto-Peto-Prentice test (generalization of the Wilcoxon rank test)

```
survdiff(Surv(time, tr.death) ~ trt, rho=1, data=pbc)
```

```
## Call:
## survdiff(formula = Surv(time, tr.death) ~ trt, data = pbc, rho = 1)
##
## n=312, 106 observations deleted due to missingness.
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=1 158     55.6     54.6    0.0183    0.0476
## trt=2 154     51.8     52.8    0.0190    0.0476
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.8
```

(f) (5) using a Weibull proportional hazards model model to estimate the hazard ratio with a 95%CI

```
summary(survreg(Surv(time, tr.death) ~ trt, dist="weibull",data=pbc))
```

```
##
## Call:
## survreg(formula = Surv(time, tr.death) ~ trt, data = pbc, dist = "weibull")
##                Value Std. Error     z      p
## (Intercept)  8.2291     0.2258 36.45 <2e-16
## trt          0.0438     0.1431  0.31  0.759
## Log(scale)  -0.1532     0.0729 -2.10  0.036
##
## Scale= 0.858
##
## Weibull distribution
## Loglik(model)= -1348.2   Loglik(intercept only)= -1348.3
##  Chisq= 0.09 on 1 degrees of freedom, p= 0.76
## Number of Newton-Raphson Iterations: 5
## n=312 (106 observations deleted due to missingness)
```

(g) (5) using a Cox's proportional hazards model to estimate the hazard ratio with a 95%CI

```
summary(coxph(Surv(time, tr.death) ~ trt, data = pbc))
```

```
## Call:
## coxph(formula = Surv(time, tr.death) ~ trt, data = pbc)
##
##   n= 312, number of events= 144
##     (106 observations deleted due to missingness)
##
##         coef exp(coef) se(coef)       z Pr(>|z|)
## trt -0.0581    0.9436   0.1669 -0.348    0.728
##
##      exp(coef) exp(-coef) lower .95 upper .95
## trt    0.9436       1.06    0.6803     1.309
##
## Concordance= 0.502  (se = 0.023 )
## Likelihood ratio test= 0.12  on 1 df,   p=0.7
## Wald test            = 0.12  on 1 df,   p=0.7
## Score (logrank) test = 0.12  on 1 df,   p=0.7
```

2.2 Lung

Use the dataset lung that comes with the library survival in R. The time and status indicators compose the endpoint.

1. (2pts) What are the levels of the status variable?

```
unique(lung$status)
```

```
## [1] 2 1
```

1=censored, 2=dead

2. (2) What proportion of patients are censored?

```
perct_censored <- lung %>%
group_by(lung$status == 1) %>%
summarise(n=n()) %>%
mutate(freq = n/sum(n))

as.numeric(perct_censored[2,3]*100)
```

```
## [1] 27.63158
```

3. (3) What is the total follow-up time in this study?

```
sum(lung$time)
```

```
## [1] 69593
```

4. (3) Calculate the mortality rate (e.g. number of deaths divided by total follow-up time)
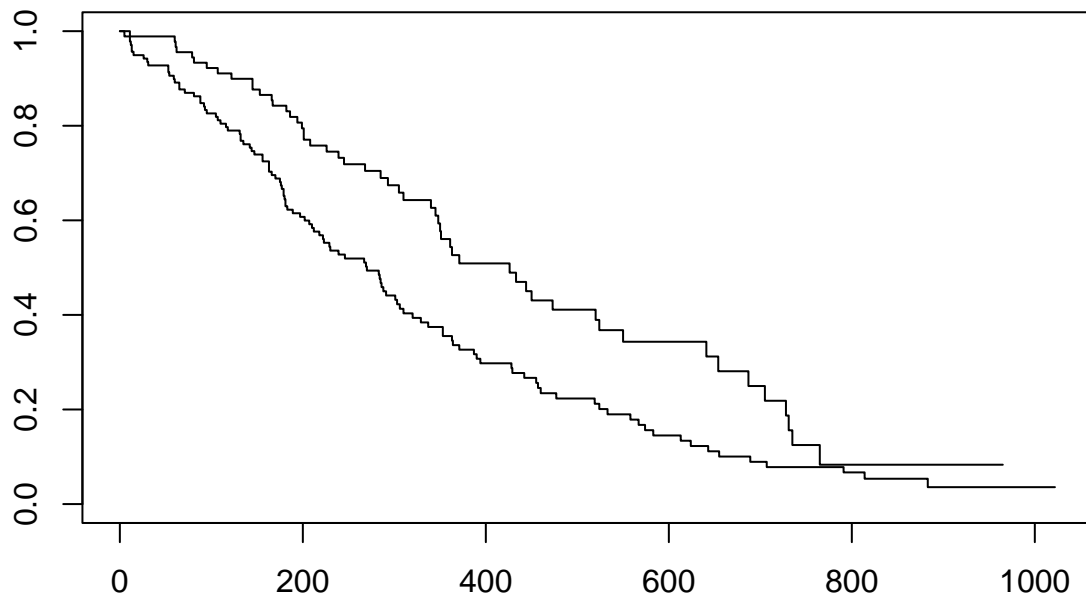
```
sum(lung$status == 2) / sum(lung$time)
```

```
## [1] 0.002370928
```

5. Examine the association of biological sex with survival as follows

(a) (3) plotting male and female survival curves

```
surv_gender <- survfit(Surv(time, status) ~ sex, data = lung)
plot(surv_gender)
```


(b) (3) reporting median survival time for each biological sex

```
male <- lung %>%
filter(lung$sex == 1) #male

female <- lung %>%
filter(lung$sex == 2) #female

print(median(male$time))
```

```
## [1] 224
```

7

```
print(median(female$time))
```

```
## [1] 292.5
```

(c) (3) conducting a log-rank test

```
survdiff(Surv(time, status) ~ sex, data = lung)
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ sex, data = lung)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 138      112     91.6      4.55      10.3
## sex=2  90       53     73.4      5.68      10.3
##
##  Chisq= 10.3  on 1 degrees of freedom, p= 0.001
```

(d) (3) conducting a Peto-Peto-Prentice test (generalization of the Wilcoxon rank test)

```
survdiff(Surv(time, status) ~ sex, rho=1, data = lung)
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ sex, data = lung, rho = 1)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 138     70.4     55.6      3.95      12.7
## sex=2  90     28.7     43.5      5.04      12.7
##
##  Chisq= 12.7  on 1 degrees of freedom, p= 4e-04
```

(e) (5) using a Weibull proportional hazards model model to estimate the hazard ratio with a 95% CI

```
summary(survreg(Surv(time, status) ~ sex, dist="weibull", data = lung))
```

```
##
## Call:
## survreg(formula = Surv(time, status) ~ sex, data = lung, dist = "weibull")
##                Value Std. Error     z       p
## (Intercept)  5.4886     0.1790 30.66 < 2e-16
## sex          0.3956     0.1276  3.10  0.0019
## Log(scale)  -0.2809     0.0619 -4.54 5.7e-06
##
## Scale= 0.755
##
## Weibull distribution
## Loglik(model)= -1148.7   Loglik(intercept only)= -1153.9
##  Chisq= 10.4 on 1 degrees of freedom, p= 0.0013
## Number of Newton-Raphson Iterations: 5
## n= 228
```

(f) (5) using a Cox's proportional hazards model to estimate the hazard ratio with a 95% CI

```r
summary(coxph(Surv(time, status) ~ sex, data = lung))
```

```
## Call:
## coxph(formula = Surv(time, status) ~ sex, data = lung)
##
##   n= 228, number of events= 165
##
##         coef exp(coef) se(coef)      z Pr(>|z|)
## sex -0.5310    0.5880   0.1672 -3.176  0.00149 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## sex     0.588      1.701    0.4237     0.816
##
## Concordance= 0.579  (se = 0.021 )
## Likelihood ratio test= 10.63  on 1 df,    p=0.001
## Wald test            = 10.09  on 1 df,    p=0.001
## Score (logrank) test = 10.33  on 1 df,    p=0.001
```

6. (2) Combine categories 2 and 3 in variable ph.ecog and report survival curves by this new variable

```r
lung$ph.ecog.females <- factor(ifelse(lung$ph.ecog >2, 2, lung$ph.ecog))
surv <- survfit(coxph(Surv(time, status) ~ ph.ecog.females, data = lung))

plot(surv)
```
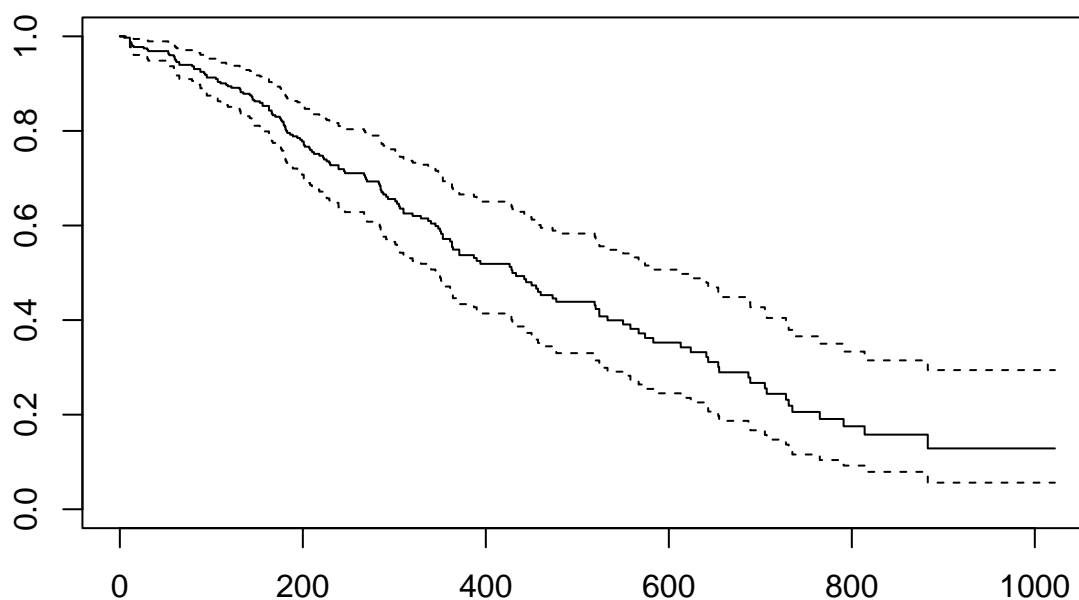
7. (3) Calculate the hazard ratios comparing each level to patients in category 1.

```
lung$ph.ecog.females <- relevel(lung$ph.ecog.females,ref="1")
hr <- coxph(Surv(time, status) ~ ph.ecog.females,data = lung)
exp(cbind(hr$coef,confint(hr)))
```

```
##                            2.5 %    97.5 %
## ph.ecog.females0 0.6919883 0.4688171 1.021396
## ph.ecog.females2 1.7554989 1.2184847 2.529188
```

8. (3) Calculate the hazard ratio for a 1 unit change in this variable (e.g. going from 1 to 2 or 2 to 3).

```
hra <- coxph(Surv(time,status) ~ ph.ecog, data=lung)
exp(cbind(hra$coef,confint(hra)))
```

```
##                    2.5 %   97.5 %
## ph.ecog 1.609532 1.288833 2.01003
```

9. (10) Examine the association of the variables age, ph.karno, pat.karno, meal.cal and wt.loss with survival using univariable Cox models and comment on the significance of the association with each.

```r
summary(oslog1 <- survreg(Surv(time, status) ~ age, dist="lognormal", data = lung))
```

```
##
## Call:
## survreg(formula = Surv(time, status) ~ age, data = lung, dist = "lognormal")
##                Value Std. Error     z      p
## (Intercept)  7.33027    0.53979 13.58 <2e-16
## age         -0.02666    0.00849 -3.14 0.0017
## Log(scale)   0.07155    0.05621  1.27 0.2030
##
## Scale= 1.07
##
## Log Normal distribution
## Loglik(model)= -1164.4   Loglik(intercept only)= -1169.3
##  Chisq= 9.83 on 1 degrees of freedom, p= 0.0017
## Number of Newton-Raphson Iterations: 3
## n= 228
```