Stats112 HW5

Viraj Vijaywargiya

2023-06-06

```
library(nnet)
library(lattice)
library(nlme)
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'lme4'
```

```
## The following object is masked from 'package:nlme':
##
##     lmList
```

```
library(geepack)
library(survival)
library(emdbook)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::collapse() masks nlme::collapse()
## x tidyr::expand()   masks Matrix::expand()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x tidyr::pack()     masks Matrix::pack()
## x tidyr::unpack()   masks Matrix::unpack()
```

1. **1a)**

```
times = c(1,1,1,6,7,7,10,10,11,12,17,17,19,23,rep(24,29))
event = c(1,1,0,1,0,0,1,1,1,1,0,0,1,1,rep(0,29))
cbind(bank=1:43,times,event)
```

```
##       bank times event
##  [1,]    1     1     1
##  [2,]    2     1     1
##  [3,]    3     1     0
##  [4,]    4     6     1
##  [5,]    5     7     0
##  [6,]    6     7     0
##  [7,]    7    10     1
##  [8,]    8    10     1
##  [9,]    9    11     1
## [10,]   10    12     1
## [11,]   11    17     0
## [12,]   12    17     0
## [13,]   13    19     1
## [14,]   14    23     1
## [15,]   15    24     0
## [16,]   16    24     0
## [17,]   17    24     0
## [18,]   18    24     0
## [19,]   19    24     0
## [20,]   20    24     0
## [21,]   21    24     0
## [22,]   22    24     0
## [23,]   23    24     0
## [24,]   24    24     0
## [25,]   25    24     0
## [26,]   26    24     0
## [27,]   27    24     0
## [28,]   28    24     0
## [29,]   29    24     0
## [30,]   30    24     0
## [31,]   31    24     0
## [32,]   32    24     0
## [33,]   33    24     0
## [34,]   34    24     0
## [35,]   35    24     0
## [36,]   36    24     0
## [37,]   37    24     0
## [38,]   38    24     0
## [39,]   39    24     0
## [40,]   40    24     0
## [41,]   41    24     0
## [42,]   42    24     0
## [43,]   43    24     0
```

1b)

j	Time Interval	nj	dj	(nj - dj)/nj	S hat (t)
0	0 <= t < 1	43	0	1	1
1	1 <= t < 6	43	2	0.953	0.953
2	6 <= t < 10	40	1	0.975	0.930
3	10 <= t < 11	37	2	0.946	0.879
4	11 <= t < 12	35	1	0.971	0.854

j	Time Interval	nj	dj	(nj - dj)/nj	S hat (t)
5	12 <= t < 19	34	1	0.971	0.829
6	19 <= t < 23	31	1	0.968	0.802
7	23 <= t <= 24	30	1	0.967	0.776
-	24 < t	-	-	-	-

1c)

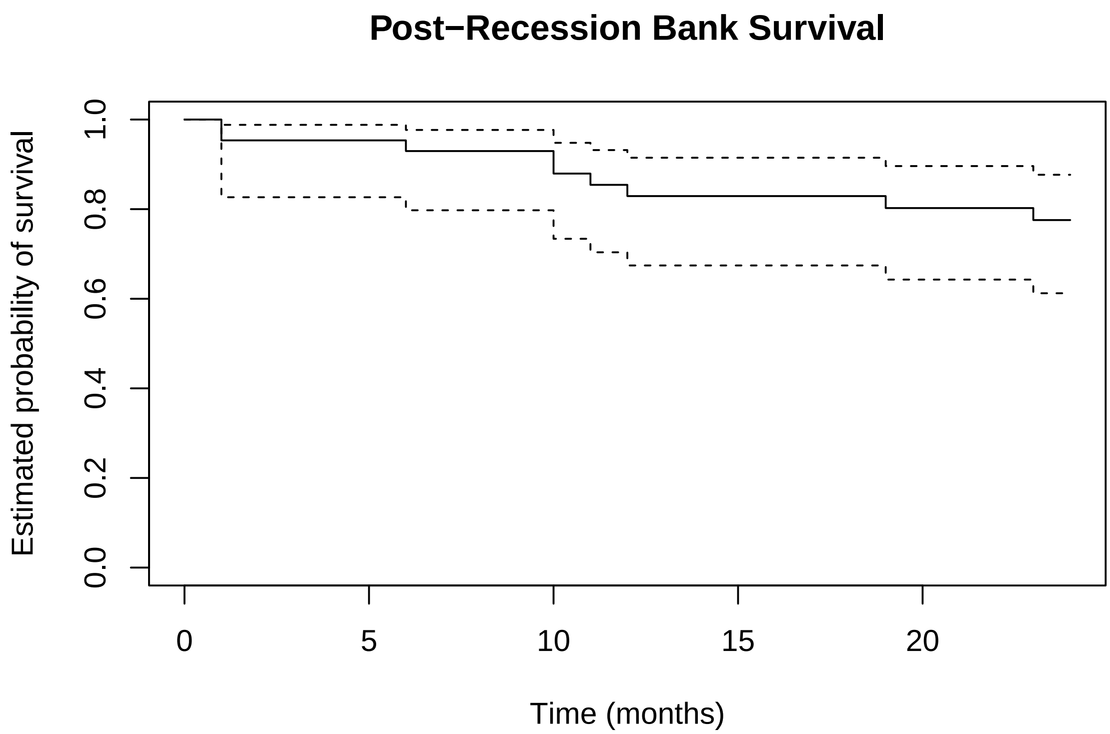
```
m1 = survfit(Surv(times,event) ~ 1, conf.type="log-log")
summary(m1)
```

```
## Call: survfit(formula = Surv(times, event) ~ 1, conf.type = "log-log")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     43       2    0.953  0.0321        0.827        0.988
##     6     40       1    0.930  0.0392        0.797        0.977
##    10     37       2    0.879  0.0507        0.734        0.948
##    11     35       1    0.854  0.0551        0.704        0.932
##    12     34       1    0.829  0.0589        0.674        0.915
##    19     31       1    0.802  0.0628        0.643        0.896
##    23     30       1    0.776  0.0662        0.612        0.877
```

80th percentile survival time: 19. 95% C.I for survival time 19 is (0.643, 0.896).
Therefore, we are 95% confident that probability of surviving past time 19 is between 0.643 and 0.896.

1d)

```
plot(m1,xlab="Time (months)",ylab="Estimated probability of survival",
     main="Post-Recession Bank Survival")
```



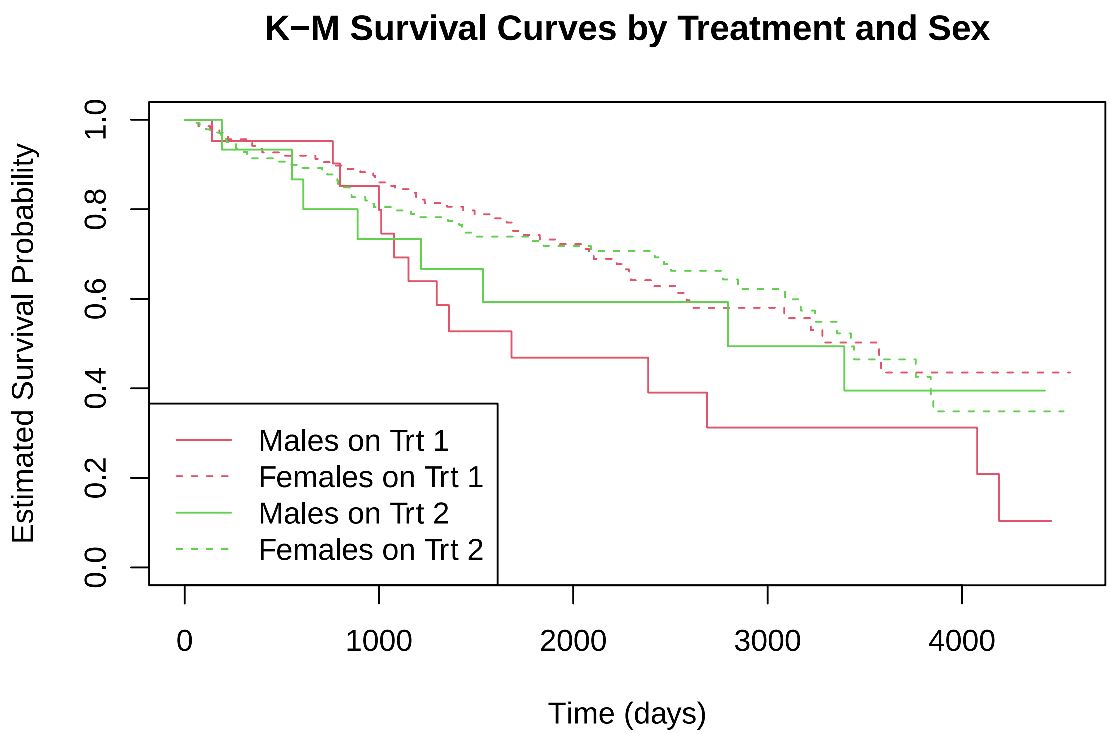
1e) Treating bank acquisitions as non-informative censoring in the above analyses may not be appropriate. In survival analysis, censoring is considered non-informative when it is unrelated to the event of interest. Non-informative censoring implies that the censoring mechanism does not depend on the unobserved event time.

In the case of bank acquisitions, treating them as non-informative censoring assumes that the probability of a bank being acquired is unrelated to its potential failure time. However, in reality, bank acquisitions may be influenced by various factors, including the financial health and performance of the bank. Banks that are more likely to fail could be more attractive targets for acquisition.

2.
```
data(pbc)
PBC = pbc[!is.na(pbc$trt),]
```

2a)

```
km = survfit(Surv(time, status==2) ~ trt+sex, data=PBC)
plot(km,col=c(2,2,3,3),lty=c(1,2,1,2),xlab="Time (days)",
     ylab="Estimated Survival Probability",
     main="K-M Survival Curves by Treatment and Sex",
     mark.time=FALSE)
legend("bottomleft",c("Males on Trt 1", "Females on Trt 1",
     "Males on Trt 2", "Females on Trt 2"),
     col=c(2,2,3,3),lty=c(1,2,1,2))
```



From the plot above, we can see that there is some overlap between the survival curves for females for each treatment. This means there seems to be no effect of treatment on the survival distribution. Whereas, for males we can see there is some difference between the survival curves comparing different treatment. Also, the treatment 2 curve seems to be a little higher than
the treatment 1 curve after around 1000 days.

2b)

```
survdiff(Surv(time, status==2) ~ trt+strata(sex), data=PBC)
```

```
## Call:
## survdiff(formula = Surv(time, status == 2) ~ trt + strata(sex),
##     data = PBC)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=1 158       65     63.6    0.0307    0.0627
## trt=2 154       60     61.4    0.0318    0.0627
##
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.8
```

H0 : h1k(t) = h2k(t) for all k = 1, 2 and all t > 0
Ha : h1k(t) = phi h2k(t) for all k = 1, 2 and all t > 0, phi != 1

p-value = 0.8, which is greater than 0.05. Therefore, we fail to reject the null and conclude that there is no evidence to reject that the hazards (and so the survival curves) differ among treatments controlling/adjusting for sex.

2c)

```r
survdiff(Surv(time, status==2) ~ trt+sex, data=PBC)
```

```
## Call:
## survdiff(formula = Surv(time, status == 2) ~ trt + sex, data = PBC)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=1, sex=m  21       14     7.73     5.080     5.447
## trt=1, sex=f 137       51    55.49     0.363     0.654
## trt=2, sex=m  15        8     6.89     0.180     0.192
## trt=2, sex=f 139       52    54.90     0.153     0.273
##
##  Chisq= 5.8  on 3 degrees of freedom, p= 0.1
```

H0 : h11(t) = h21(t) = h12(t) = h22(t).

This hypothesis has a stronger condition than the hypothesis in the prev part b because it assumes all four survival distributions are the same.

2d)

```r
PBC$log.bili = log(PBC$bili)
mod.d = coxph(Surv(time,status==2)~log.bili,data=PBC)
summary(mod.d)
```

```
## Call:
## coxph(formula = Surv(time, status == 2) ~ log.bili, data = PBC)
##
##   n= 312, number of events= 125
##
##              coef exp(coef) se(coef)     z Pr(>|z|)
## log.bili 1.08524   2.96016  0.09333 11.63   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## log.bili      2.96     0.3378     2.465     3.554
##
## Concordance= 0.794  (se = 0.02 )
## Likelihood ratio test= 133.2  on 1 df,   p=<2e-16
## Wald test            = 135.2  on 1 df,   p=<2e-16
## Score (logrank) test = 159.2  on 1 df,   p=<2e-16
```

Fitted model: hi hat(t) = exp(1.085 log(xi)) h0 hat(t); where hi(t) is the hazard rate (of death) for individual i, xi is the bilirubin level (mg/dl) for subject i, and h0(t) is the hazard function when log(xi) = 0.

For a 1 unit increase in bilirubin level on the log scale (log(bili) increase 1 unit), the estimated hazard rate of death increases by 196%. A 1 unit increase in log(bili) is estimated to have 2.96 times the hazard of the original log(bili) level.

2e)

```r
mod.e = coxph(Surv(time,status==2)~log(bili)*factor(edema)+age,data=PBC)
summary(mod.e)
```

```
## Call:
## coxph(formula = Surv(time, status == 2) ~ log(bili) * factor(edema) +
##     age, data = PBC)
##
##   n= 312, number of events= 125
##
##                                  coef exp(coef)  se(coef)      z Pr(>|z|)
## log(bili)                    0.992647  2.698367  0.110389  8.992  < 2e-16 ***
## factor(edema)0.5             0.064943  1.067098  0.449646  0.144   0.8852
## factor(edema)1               2.489309 12.052942  0.518478  4.801 1.58e-06 ***
## age                          0.042426  1.043338  0.008552  4.961 7.03e-07 ***
## log(bili):factor(edema)0.5  0.196390  1.217001  0.227633  0.863   0.3883
## log(bili):factor(edema)1   -0.497822  0.607853  0.266848 -1.866   0.0621 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                            exp(coef) exp(-coef) lower .95 upper .95
## log(bili)                     2.6984    0.37059    2.1734     3.350
## factor(edema)0.5              1.0671    0.93712    0.4420     2.576
## factor(edema)1               12.0529    0.08297    4.3628    33.298
## age                           1.0433    0.95846    1.0260     1.061
## log(bili):factor(edema)0.5    1.2170    0.82169    0.7790     1.901
## log(bili):factor(edema)1      0.6079    1.64513    0.3603     1.026
##
## Concordance= 0.839  (se = 0.019 )
## Likelihood ratio test= 182.2  on 6 df,   p=<2e-16
## Wald test            = 190.8  on 6 df,   p=<2e-16
## Score (logrank) test = 292.6  on 6 df,   p=<2e-16
```

Fitted hazard ratio for the ith individual:
(hi(t)/h0(t)) hat = exp{0.993 log(xi) + 0.0649 E0.5i + 2.489 E1i + 0.0424 ai
0.196 log(xi) * E0.5i - 0.498 log(xi) * E1i}; where xi is the bilirubin level for the ith individual, and ai is the age (years) of the ith individual.
E0.5i: 1 = edema untreated or successfully treated, 0 = else.
E1i: 1 = edema despite diuretic therapy, 0 = else.

h0(t) is the hazard rate for an individual with no edema, bilirubin level 1 mg/dl, and age zero.

INTERPRETATIONS
log(bili): For individuals with no edema and holding age constant, for each 1 unit increase in log bilirubin level, the estimated hazard rate increases by about 170 (exp(.993) = 2.70).
log(bili):factor(edema)1: For a 1 unit increase in log bilirubin level, the estimated increase in hazard rate for individuals with no edema is 170% (exp(0.992) = 2.7), but for individuals with edema despite diuretic therapy, the estimated increase is only about 64% (exp(.9926 - .4978) = 1.64), holding age constant.

age: The estimated hazard rate increases by about 4% (exp(.0424) = 1.04), for each year increase in age, holding bilirubin level and edema status constant.

2f)

```
anova(mod.d, mod.e)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(time, status == 2)
##  Model 1: ~ log.bili
```

```
##  Model 2: ~ log(bili) * factor(edema) + age
##    loglik  Chisq Df Pr(>|Chi|)
## 1 -573.39
## 2 -548.85 49.089  5  2.128e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H0: reduced model (mod.d) is better. Ha: full model (mod.e) is better.
p-value = 2.128e-09. Therefore, we reject the null (reduced model) and conclude that there is significant evidence that, adjusting for (log) bilirubin level, age, edema status, and the interaction between edema status and (log) bilirubin level help predict the hazard rate of death in the population of patients with primary biliary cirrhosis of the liver.