Stats 112 Homework 5

Owen Lin

Problem 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:

```
Time_interval = c(1, 6, 10, 11, 12, 19, 23)
nj = c(43, 40, 37, 35, 34, 31, 30)
dj = c(2, 1, 2, 1, 1, 1, 1)
nj_dj_divides_nj = (nj-dj)/nj
S_t = rep(0, 7)
for (i in 1:7){
    S_t[i] = prod(nj_dj_divides_nj[1:i])
}
round(cbind(Time_interval, nj, dj, nj_dj_divides_nj, S_t),3)
```

```
##      Time_interval nj dj nj_dj_divides_nj   S_t
## [1,]             1 43  2            0.953 0.953
## [2,]             6 40  1            0.975 0.930
## [3,]            10 37  2            0.946 0.879
## [4,]            11 35  1            0.971 0.854
## [5,]            12 34  1            0.971 0.829
## [6,]            19 31  1            0.968 0.802
## [7,]            23 30  1            0.967 0.776
```

1c:

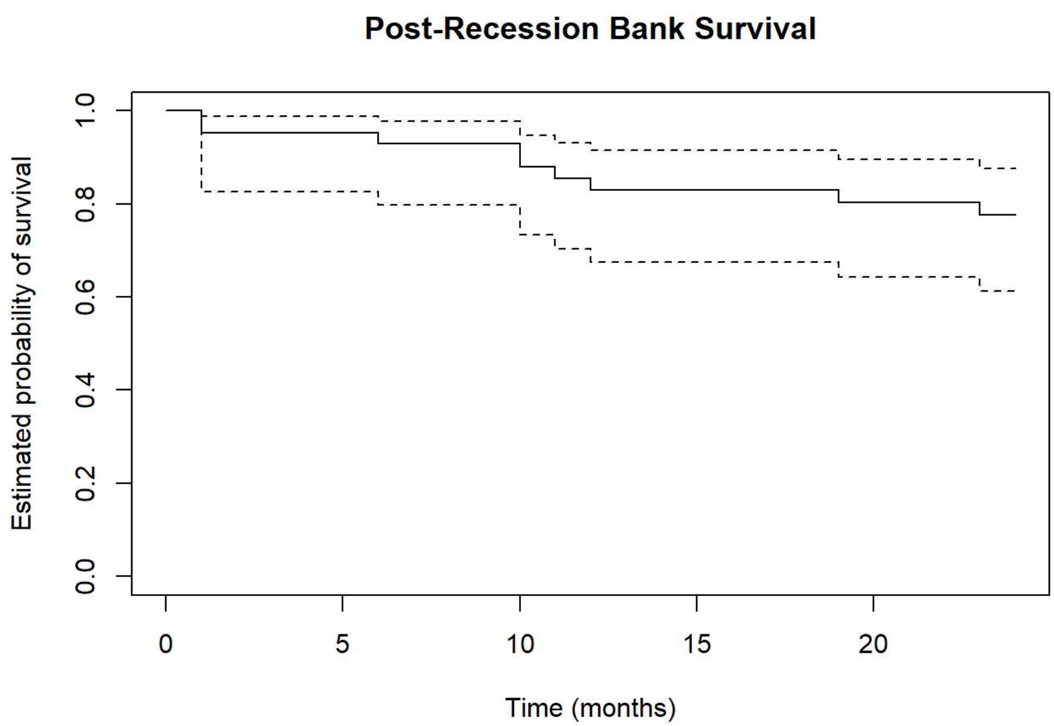
The 80-th percentile survival time is at month 19.
The 95% confidence interval for this estimate is (0.643, 0.896).
We are 95% confident that about 64.3% to 89.6% of the banks will not fail before month 19.

```
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
```

1d:

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



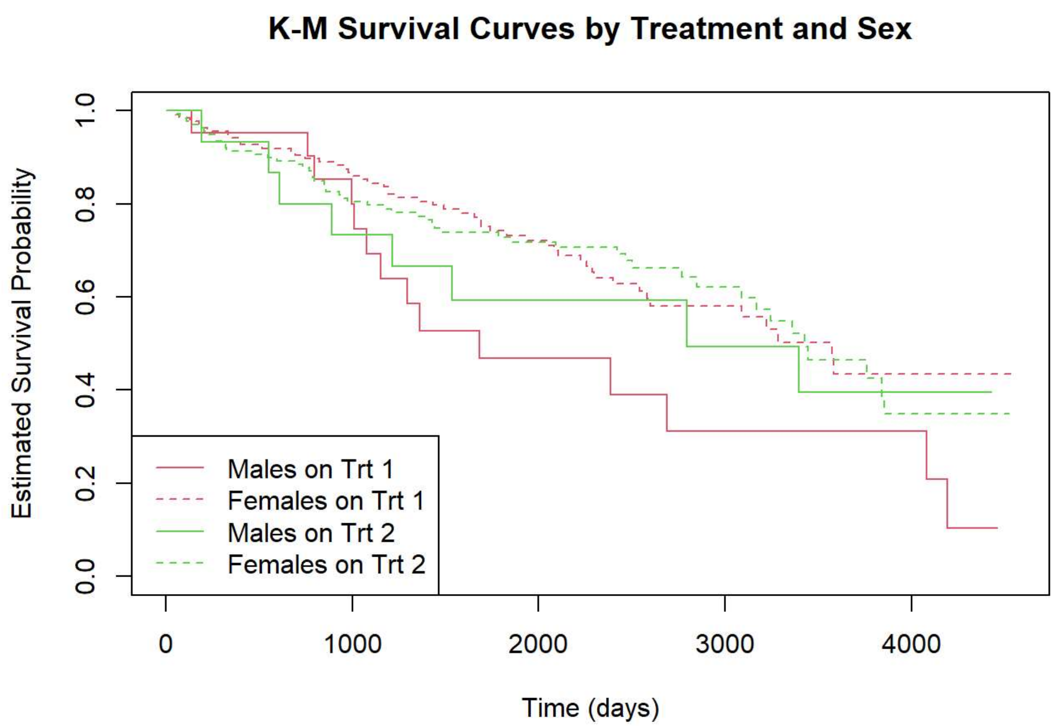
1e:
It’s not appropriate to treat the bank acquisitions as non-informative censoring because it is reasonable to consider that the bank was acquired when it was on the verge of bankruptcy.

Problem 2

2a:
In general, treatment2 (The placebo group) has a higher probability of survival, and females seem to have a higher probability of survival than males.

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

km = survfit(Surv(time, status==2) ~ trt+sex, conf.type="log-log", 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))
```



2b:
H0: S_i(t) = S_i(t) for i = treatment groups
Ha: H0 is not true
test statistic: 0.0627 ~ chi_squared(1)
p-value: 0.8
decision: fail to reject the null
conclusion: the survival time between the two treatment groups are the same for all time.

```
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
```

2c:
H0: S_ij(t) = S_ij(t) for all i, j = 1, 2, (group and sex)
2c treats it as 4 different groups, and 2b treats it as 2 groups stratified by sex.

```
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
```

2d:

$$\ln \frac{h_i(t)}{h_0(t)} = 1.08 * \ln bili$$

where bili is the serum bilirunbin in mg/dl
1 percentage change in serum bilirunbin will result in a relative change in hazard ratios of $e^{1.08}$

```
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
```

2e:

$$\frac{h_i(t)}{h_0(t)} = 2.69 * \ln bili + 1.06 * edema0.5 + 12.05 * edema1 + 1.04 * age + 1.21 * \ln bili * edema0.5 + 0.60 * \ln bili * edema1$$

where bili is the serum bilirunbin in mg/dl, age is in years, edema0.5 is untreated or successfully treated, and edema1 being edema despite diuretic therapy.
If no edema, 1 percentage change in serum bilirunbin will result in a relative change in hazard ratios of 2.69.
For the group with edema despite diuretic therapy, 1 percentage change in serum bilirunbin will result in an additional relative change in hazard ratios of 1.21.
1 year increases in age will result in a relative change in hazard ratios of 1.04.

```
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
```

2f:

The p-value is 2.128*10^-9, so we reject the simpler model and conclude that edema and age are significant factors in hazard ratios.

```
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
```