## 2

```r
library(survival)
time = c(3,6,7,7,10,10,10,10,13,
         14,15,15,19,19,20,20,22,22,
         27,27,28,29,30,31,34,35,36,
         37,38,39,39,39,39,40,42,43,
         43,47,47,50,58,59,70,74,78)
status = c(1,1,0,1,1,1,0,0,1,
           1,1,0,1,1,0,1,1,1,
           1,1,1,0,1,0,1,1,1,
           1,1,1,1,0,1,1,1,1,
           0,1,1,0,0,0,0,0,0)
fit <- survfit(Surv(time,status) ~ 1, type="kaplan-meier")
```

### a

Calculate the estimate of the hazard probability at time t = 47.

```r
summary(fit,time = 47)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, type = "kaplan-meier")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    47      8      30    0.204  0.0697        0.105        0.399
```

Number at risk: 8

Events: 2

$\hat{\lambda}(47) = 2/8 = 0.25$

### b

```r
summary(fit,time = c(3,6,7,10,13))
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, type = "kaplan-meier")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     3     45       1    0.978  0.0220        0.936        1.000
##     6     44       1    0.956  0.0307        0.897        1.000
##     7     43       1    0.933  0.0372        0.863        1.000
##    10     41       2    0.888  0.0473        0.800        0.986
##    13     37       1    0.864  0.0517        0.768        0.971
```

Estimate the survival function at time t = 13

$$\hat{S}(13) = (1 - \hat{\lambda}_3)(1 - \hat{\lambda}_6)(1 - \hat{\lambda}_7)(1 - \hat{\lambda}_{10})(1 - \hat{\lambda}_{13})$$

$$= (1 - 1/45)(1 - 1/44)(1 - 1/43)(1 - 2/41)(1 - 1/37) = 0.8638$$

### c

Estimate the 20% percentile of the distribution of the service times.

```
unlist(quantile(fit,0.2))
```

```
## quantile.20    lower.20    upper.20
##          19          10          28
```

$$P(T \leq 19) = 0.2$$

The 20% percentile of the distribution of the service times is 19.

## Q3

The data set Neph.txt contains observation from a small kidney disease study. The first column in the data file is an indicator of whether or not the subjects received a nephrectomy treatment. The second column is a factor dividing the subjects into three age ranges. The last two columns measure the time until an event and the status. We wish to test whether the nephrectomy treatment extended the times until failure.

```
neph <- read.table("Neph.txt", header=TRUE,
                   colClasses = c("factor","factor","numeric","numeric"))
```

### a

Estimate the median survival time and give a 95% confidence interval.

```
fit <- survfit(Surv(time,status) ~ 1, type="kaplan-meier",data = neph)
unlist(quantile(fit,0.5))
```

```
## quantile.50    lower.50    upper.50
##          26          15          52
```

Estimated median: 26

95% confidence interval: (15,52)

### b

Perform a likelihood ratio test for an effect from the nephrectomy treatment when we control for the differences between the ages using a Cox PH model. Report a P value for the test.

```
fit <- coxph(Surv(time,status)~
     age+nephrectomy,data=neph)
anova(fit)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(time, status)
## Terms added sequentially (first to last)
##
##               loglik  Chisq Df Pr(>|Chi|)
## NULL        -88.293
## age         -85.400 5.7857  2    0.05542 .
## nephrectomy -82.112 6.5773  1    0.01033 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p value is $0.01033 < 0.05$.

Conclusion: reject the null so there is significant effect from the nephrectomy treatment.

### c

Repeat this test except use a model that is stratified on the age groups

```r
fit <- coxph(Surv(time,status)~
   strata(age)+nephrectomy,data=neph)
anova(fit)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(time, status)
## Terms added sequentially (first to last)
##
##               loglik  Chisq Df Pr(>|Chi|)
## NULL        -58.857
## nephrectomy -56.147 5.4204  1     0.0199 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p value is $0.0199 < 0.05$.

Conclusion: reject the null so there is significant effect from the nephrectomy treatment.

### d

Describe how the two models in (b) and (c) are different, and what factors you would consider when deciding which one is the correct test to use.

The difference between (b) and (c) is that in (b) the variable age is not stratified, so the cox ph model can estimate the slope of the age. In model (c), age is stratified then the cox ph model can no longer estimate the age effect.

When the age effect is proportional, then we should use the model (b) because we could have parametric estimation on the age.

When the age effect is not proportional, we cannot use the model (b), and we have to stratify the age effect. Otherwise the test on the treatment may be wrong.