## Q2

```
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(1,1,0,0,1,1,1,0,1,
           1,1,0,1,1,0,1,1,1,
           1,1,1,0,1,0,1,1,1,
           1,1,1,1,1,1,1,1,1,
           1,1,0,0,0,0,0,0,0)
sur = Surv(time,status)
KM <- survfit(sur ~ 1, type="kaplan-meier")
summary(KM)
```

```
## Call: survfit(formula = sur ~ 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
##    10     41       3    0.886  0.0482        0.796        0.985
##    13     37       1    0.862  0.0525        0.765        0.971
##    14     36       1    0.838  0.0562        0.735        0.956
##    15     35       1    0.814  0.0595        0.705        0.939
##    19     33       2    0.765  0.0653        0.647        0.904
##    20     31       1    0.740  0.0677        0.618        0.885
##    22     29       2    0.689  0.0720        0.561        0.845
##    27     27       2    0.638  0.0752        0.506        0.804
##    28     25       1    0.612  0.0764        0.479        0.782
##    30     23       1    0.586  0.0776        0.452        0.759
##    34     21       1    0.558  0.0787        0.423        0.735
##    35     20       1    0.530  0.0796        0.395        0.711
##    36     19       1    0.502  0.0801        0.367        0.686
##    37     18       1    0.474  0.0804        0.340        0.661
##    38     17       1    0.446  0.0803        0.314        0.635
##    39     16       4    0.335  0.0772        0.213        0.526
##    40     12       1    0.307  0.0757        0.189        0.497
##    42     11       1    0.279  0.0737        0.166        0.468
##    43     10       2    0.223  0.0687        0.122        0.408
##    47      8       1    0.195  0.0656        0.101        0.377
```

#### a

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

$\hat{h}_{47} = 1/8 = 0.125$

#### b

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

$$= (1 - 1/45)(1 - 1/44)(1 - 3/41)(1 - 1/37) = 0.8617007$$

The result is the same with the R calculation result (0.862).

```
summary(KM,t = 13)
```

```
## Call: survfit(formula = sur ~ 1, type = "kaplan-meier")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    13     37       6    0.862  0.0525        0.765        0.971
```

**c**

The 20% percentile of the distribution of the service times is the first observation where the estimated survival drops below 80%.

```
unlist(quantile(KM,0.2))
```

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

In the R table, we can see this happens at 19. Therefore $P(T \leq 19) = 0.2$

## Q3

```
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.

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

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

So the estimated median is 26 and the 95% confidence interval is (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.

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

```
## 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
```

We control age by make the sequential order age first.

So the p value for the likelihood ratio test is $0.01033 < 0.05$.

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

```
## 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 for the likelihood ratio test is $0.0199 < 0.05$.

There is significant effect from the nephrectomy treatment.

But the p value is bigger than the part b.

**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 model in (b) is the cox proportion hazard model and treat age as a predictor, and the model in (c) is a stratified model which treat the age as a confounding variable that do not satisfy the proportional hazard assumption.

If the age satisfies the proportional hazard assumption and we are interested in the effect in age, we should use the model in (b), because it can provide parametric estimation and confidence interval on age.

If the age does not satisfy the proportional hazard assumption, or we are not interested in the effect of age, we should use the stratified model.