

## Chapter 11 Survival Analysis

```
library(ISLR2)
library(survival)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

### Exercise 1

Independent: a, d, e, g, i

Not independent: b, c, f, h

### Exercise 2

	1	2	3	4
c	unknown	2	1.5	unknown
k	1.2	unknown	unknown	0.2
y	1.2	2	1.5	0.2
$\delta$	1	0	0	1

### Exercise 3

t	$r_j$	$d_j$	$q_j$
0	4	0	0
0.2	4	1	0
1.2	3	1	1
2	1	0	1

## Exercise 4

a

$$S(50) = 0.6$$

b

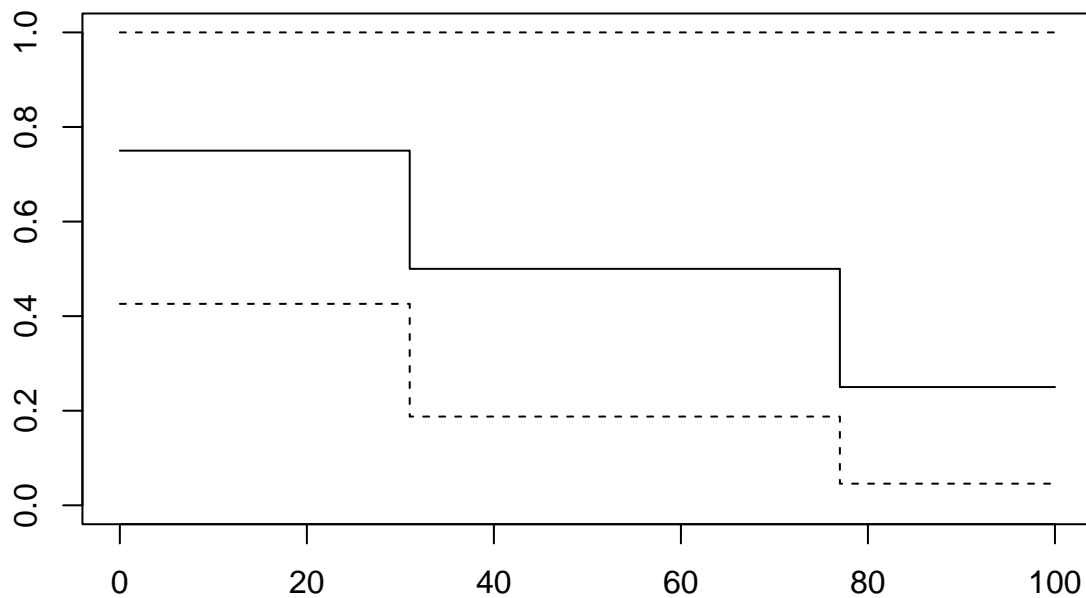
$$S(t) = 0.8 \text{ if } t < 37.2$$

$$S(t) = 0.6 \text{ if } t < 57.3$$

$$S(t) = 0.4 \text{ if } t \geq 57.3$$

## Exercise 5

```
data = data.frame(time = c(0, 31, 77, 100), status = c(1, 1, 1, 0))  
data_curve = survfit(Surv(time, status) ~ 1, data = data)  
plot(data_curve)
```



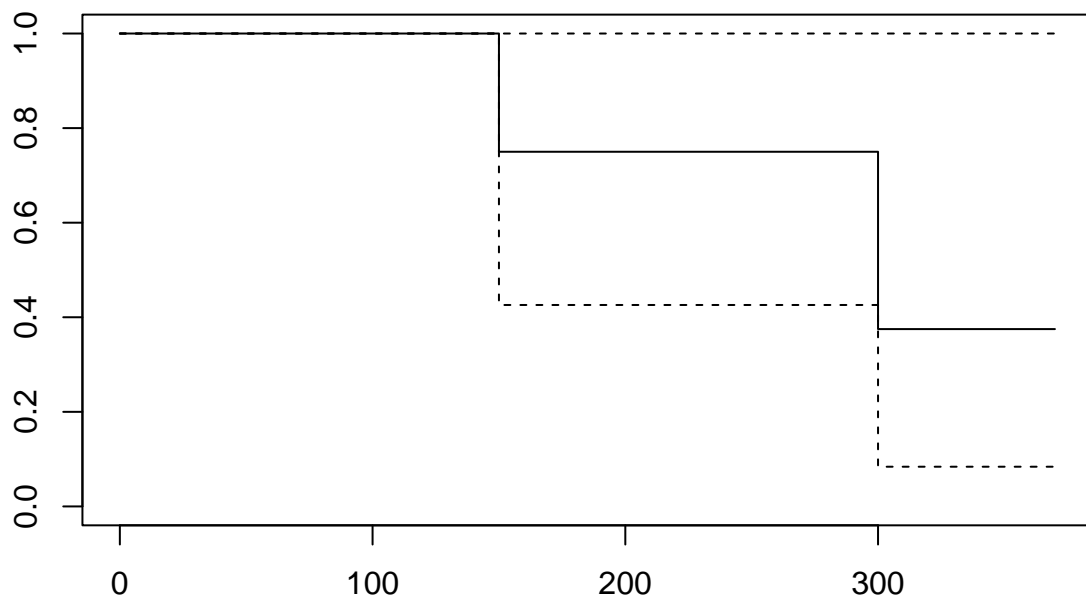
## Exercise 6

a

$t$	$\delta_j$	$r_j$	$d_j$	$q_j$
0	1	4	0	0
150	0	4	1	0
300	1	3	1	1
370	0	1	0	1

b

```
data = data.frame(time = c(150, 250, 300, 370), status = c(1, 0, 1, 0))
data_curve = survfit(Surv(time, status) ~ 1, data = data)
plot(data_curve)
```



```
summary(data_curve)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, data = data)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   150      4      1    0.750   0.217    0.4259          1
##   300      2      1    0.375   0.286    0.0839          1
```

**c**

$$S(t) = 0.15$$

$$S(t) = 0.375$$

**d**

$$S(t) = 0.75 \text{ if } t < 300$$

$$S(t) = 0.375 \text{ if } t$$

$$\geq$$

$$300$$

## Exercise 7

**a**

If we think  $q$  (the number of deaths) as “successes” (although we don’t mean it literally),  $q$  follows Hypergeometric distribution

$$q_1 \sim Hyper(r_k, q_k, r_{1k})$$

$$P(q_1 = q'_{1k}) = \frac{\binom{q_k}{q'_{1k}} \binom{r_k - q_k}{r_{1k} - q'_{1k}}}{\binom{r_k}{r_{1k}}}$$

where

$$r_k$$

is the population size,

$$q_k$$

is the number of successes,

$$r_{1k}$$

is the number of draws,

$$q'_{1k}$$

is the number of observed successes

**b**

$$Mean(q_{1k}) = r_{1k} \cdot \frac{q_k}{r_k}$$

$$Variance(q_{1k}) = r_{1k} \cdot \frac{q_k}{r_k} \cdot \frac{r_k - q_k}{r_k} \cdot \frac{r_k - r_{1k}}{r_k - 1}$$

## Exercise 8

a

$$\begin{aligned}
 &= \lim_{\Delta t \rightarrow 0} = \frac{F(t + \Delta t) - F(t)}{\Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} = \frac{1 - S(t + \Delta t) - F(t)}{\Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} = \frac{1 - S(t + \Delta t) - P(T < t)}{\Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} = \frac{1 - P(T > t + \Delta t) - P(T < t)}{\Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} = \frac{P(T \leq t + \Delta t) - P(T < t)}{\Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} = \frac{P(t < T \leq t + \Delta t)}{\Delta t} \\
 &= f(t)
 \end{aligned}$$

b

$$\begin{aligned}
 &\frac{dF(t)}{dt} = f(t) \\
 &\Leftrightarrow -\frac{dS(t)}{dt} = f(t) \\
 &\Leftrightarrow -\frac{dS(t)}{dt} \frac{1}{S(t)} = \frac{f(t)}{S(t)} \\
 &\Leftrightarrow -\frac{d \log(S(t))}{dt} = h(t) \\
 &\Leftrightarrow -\log(S(t)) = \int_0^t h(u) du \\
 &\Leftrightarrow S(t) = \exp \left( - \int_0^t h(u) du \right)
 \end{aligned}$$

## Exercise 9

a

$$T \sim \text{Exp}(\lambda)$$

$$\Rightarrow P(T = t) = \lambda e^{-\lambda t}$$

$$\Rightarrow P(T > t) = e^{-\lambda t}$$

$$\Rightarrow S(t) = e^{-\lambda t}$$

b

$$X \sim \text{Exp}(\lambda)$$

$$\Rightarrow \hat{\lambda}_{MLE} = \frac{1}{\bar{x}} = \frac{n}{\sum_i x_i}$$

We have

$$T \sim \text{Exp}(\lambda) \text{ and } t_1, t_2, \dots, t_n$$

are independent samples from T (t could be the true survival time or the censoring time and we also assume that censoring mechanism is independent). So,

$$\lambda_{MLE} = \frac{n}{\sum_i t_i} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n y_i}$$

And,

$$\text{Mean}(T) = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n \delta_i}$$

## Exercise 10

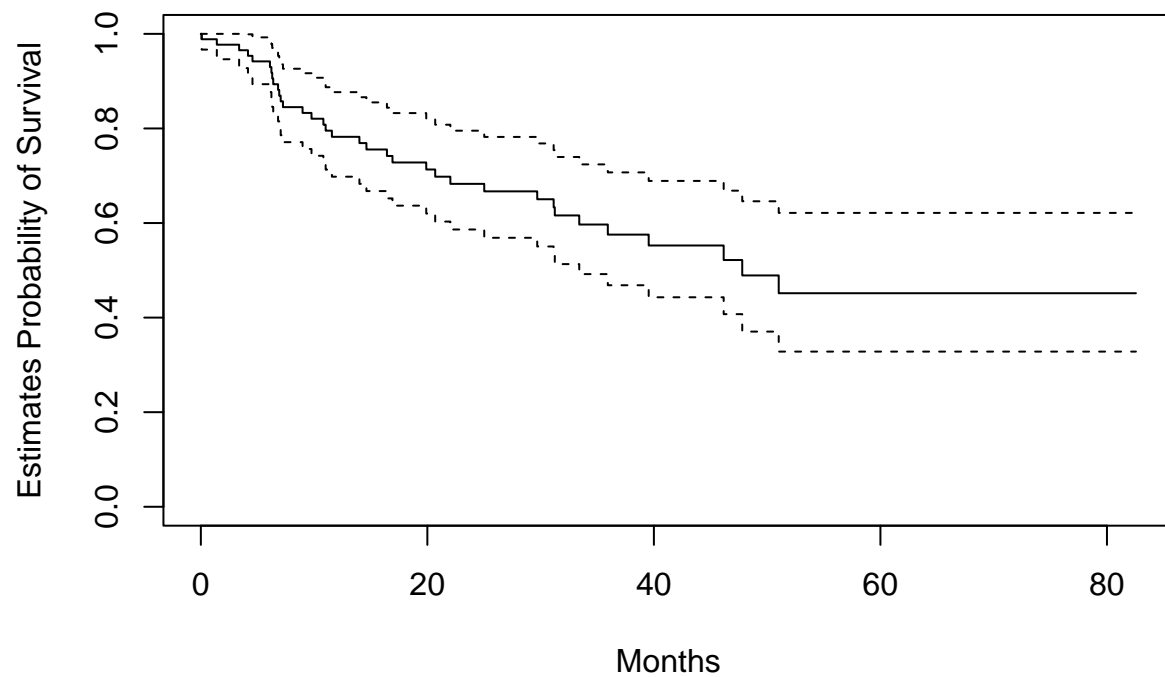
```
library(ISLR2)
library(survival)
library(dplyr)
```

```
head(BrainCancer)
```

```
##      sex  diagnosis      loc ki   gtv stereo status  time
## 1 Female Meningioma Infratentorial 90 6.11    SRS      0 57.64
## 2 Male   HG glioma Supratentorial 90 19.35    SRT      1  8.98
## 3 Female Meningioma Infratentorial 70 7.95    SRS      0 26.46
## 4 Female LG glioma Supratentorial 80 7.61    SRT      1 47.80
## 5 Male   HG glioma Supratentorial 90 5.06    SRT      1  6.30
## 6 Female Meningioma Supratentorial 80 4.82    SRS      0 52.75
```

a

```
attach(BrainCancer)
km_95_curve = survfit(Surv(time, status) ~ 1, conf.int = 0.95)
plot(km_95_curve, xlab = 'Months', ylab = 'Estimates Probability of Survival')
```

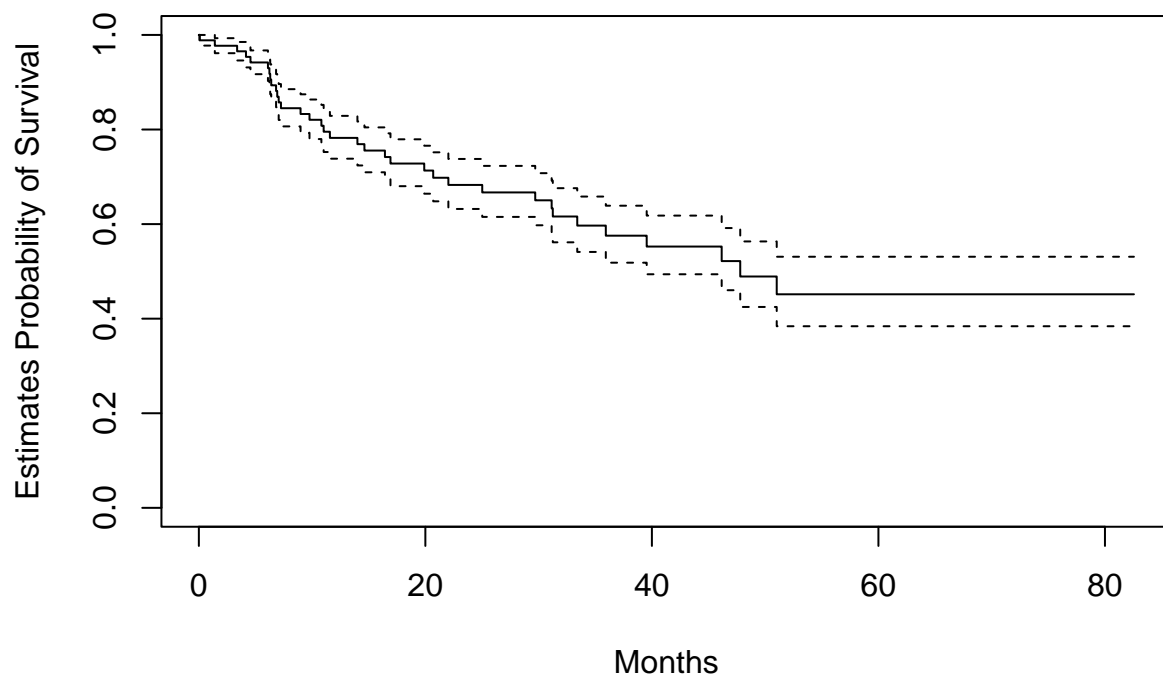


```
summary(km_95_curve)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, conf.int = 0.95)
##
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  0.07      88       1   0.989  0.0113   0.967    1.000
##  1.41      86       1   0.977  0.0160   0.946    1.000
##  3.38      83       1   0.965  0.0197   0.928    1.000
##  4.16      82       1   0.954  0.0227   0.910    0.999
##  4.56      81       1   0.942  0.0253   0.894    0.993
##  6.10      78       1   0.930  0.0277   0.877    0.986
##  6.23      77       1   0.918  0.0298   0.861    0.978
##  6.30      76       1   0.906  0.0318   0.845    0.970
##  6.39      75       1   0.894  0.0336   0.830    0.962
##  6.82      74       1   0.881  0.0352   0.815    0.953
##  6.92      73       1   0.869  0.0368   0.800    0.944
##  7.05      72       1   0.857  0.0382   0.786    0.935
##  7.25      70       1   0.845  0.0395   0.771    0.926
```

##	8.98	69	1	0.833	0.0408	0.757	0.917
##	9.77	68	1	0.821	0.0420	0.742	0.907
##	10.82	65	1	0.808	0.0432	0.727	0.897
##	11.02	64	1	0.795	0.0444	0.713	0.887
##	11.57	61	1	0.782	0.0455	0.698	0.877
##	14.00	59	1	0.769	0.0466	0.683	0.866
##	14.62	57	1	0.756	0.0477	0.668	0.855
##	16.43	55	1	0.742	0.0488	0.652	0.844
##	16.92	54	1	0.728	0.0498	0.637	0.832
##	19.90	49	1	0.713	0.0509	0.620	0.820
##	20.69	47	1	0.698	0.0521	0.603	0.808
##	22.03	46	1	0.683	0.0531	0.586	0.795
##	25.02	43	1	0.667	0.0542	0.569	0.782
##	29.70	40	1	0.650	0.0553	0.550	0.768
##	31.15	38	1	0.633	0.0565	0.532	0.754
##	31.25	37	1	0.616	0.0575	0.513	0.740
##	33.41	32	1	0.597	0.0588	0.492	0.724
##	35.93	28	1	0.575	0.0605	0.468	0.707
##	39.54	25	1	0.552	0.0623	0.443	0.689
##	46.16	18	1	0.522	0.0659	0.407	0.668
##	47.80	16	1	0.489	0.0694	0.370	0.646
##	51.02	13	1	0.452	0.0736	0.328	0.621

```
km_68_curve = survfit(Surv(time, status) ~ 1, conf.int = 0.68)
plot(km_68_curve, xlab = 'Months', ylab = 'Estimates Probability of Survival')
```





```
summary(km_68_curve)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, conf.int = 0.68)
##
##   time n.risk n.event survival std.err lower 68% CI upper 68% CI
##   0.07    88      1   0.989  0.0113   0.977    1.000
##   1.41    86      1   0.977  0.0160   0.961    0.993
##   3.38    83      1   0.965  0.0197   0.946    0.985
##   4.16    82      1   0.954  0.0227   0.931    0.976
##   4.56    81      1   0.942  0.0253   0.917    0.967
##   6.10    78      1   0.930  0.0277   0.903    0.958
##   6.23    77      1   0.918  0.0298   0.888    0.948
##   6.30    76      1   0.906  0.0318   0.875    0.938
##   6.39    75      1   0.894  0.0336   0.861    0.928
##   6.82    74      1   0.881  0.0352   0.847    0.917
##   6.92    73      1   0.869  0.0368   0.834    0.907
##   7.05    72      1   0.857  0.0382   0.820    0.896
##   7.25    70      1   0.845  0.0395   0.807    0.885
##   8.98    69      1   0.833  0.0408   0.793    0.874
##   9.77    68      1   0.821  0.0420   0.780    0.863
##  10.82    65      1   0.808  0.0432   0.766    0.852
##  11.02    64      1   0.795  0.0444   0.752    0.841
##  11.57    61      1   0.782  0.0455   0.738    0.829
##  14.00    59      1   0.769  0.0466   0.724    0.817
##  14.62    57      1   0.756  0.0477   0.710    0.805
##  16.43    55      1   0.742  0.0488   0.695    0.792
##  16.92    54      1   0.728  0.0498   0.680    0.779
##  19.90    49      1   0.713  0.0509   0.664    0.766
##  20.69    47      1   0.698  0.0521   0.648    0.752
##  22.03    46      1   0.683  0.0531   0.632    0.738
##  25.02    43      1   0.667  0.0542   0.615    0.723
##  29.70    40      1   0.650  0.0553   0.598    0.708
##  31.15    38      1   0.633  0.0565   0.579    0.692
##  31.25    37      1   0.616  0.0575   0.561    0.676
##  33.41    32      1   0.597  0.0588   0.541    0.658
##  35.93    28      1   0.575  0.0605   0.518    0.639
##  39.54    25      1   0.552  0.0623   0.494    0.618
##  46.16    18      1   0.522  0.0659   0.460    0.592
##  47.80    16      1   0.489  0.0694   0.425    0.563
##  51.02    13      1   0.452  0.0736   0.384    0.531
```

b

```
# an example
braincancer_by_time = data.frame(time = sort(BrainCancer$time))
set.seed(42)

bootstrap_indices = sample(nrow(BrainCancer), size = 88, replace = T)
bootstrap_sample = BrainCancer[bootstrap_indices, ]
bootstrap_curve = survfit(Surv(time, status) ~ 1, data = bootstrap_sample)
```

```

#plot(bootstrap_curve)
#summary(bootstrap_curve)

bootstrap_df_sample = data.frame(time = bootstrap_curve$time,
                                  survive = bootstrap_curve$surv)
braincancer_by_time

```

```

##      time
## 1  0.07
## 2  1.18
## 3  1.41
## 4  1.54
## 5  2.03
## 6  3.38
## 7  4.16
## 8  4.56
## 9  5.15
## 10 5.51
## 11 6.10
## 12 6.23
## 13 6.30
## 14 6.39
## 15 6.82
## 16 6.92
## 17 7.05
## 18 7.18
## 19 7.25
## 20 8.98
## 21 9.77
## 22 10.10
## 23 10.49
## 24 10.82
## 25 11.02
## 26 11.48
## 27 11.51
## 28 11.57
## 29 13.90
## 30 14.00
## 31 14.56
## 32 14.62
## 33 14.75
## 34 16.43
## 35 16.92
## 36 17.57
## 37 18.95
## 38 19.41
## 39 19.74
## 40 19.90
## 41 20.13
## 42 20.69
## 43 22.03
## 44 23.67
## 45 24.39

```

```
## 46 25.02
## 47 26.46
## 48 26.85
## 49 29.70
## 50 30.16
## 51 31.15
## 52 31.25
## 53 31.67
## 54 32.82
## 55 32.82
## 56 32.98
## 57 33.41
## 58 33.67
## 59 34.26
## 60 34.66
## 61 35.93
## 62 36.10
## 63 39.25
## 64 39.54
## 65 39.54
## 66 41.44
## 67 42.07
## 68 42.10
## 69 44.39
## 70 45.74
## 71 46.16
## 72 47.11
## 73 47.80
## 74 49.05
## 75 50.85
## 76 51.02
## 77 52.23
## 78 52.75
## 79 54.43
## 80 55.80
## 81 57.11
## 82 57.25
## 83 57.64
## 84 65.02
## 85 67.38
## 86 73.74
## 87 78.75
## 88 82.56
```

```
arrange(bootstrap_sample, time)
```

##	sex	diagnosis	loc	ki	gtv	stereo	status	time
## 83	Male	LG glioma	Infratentorial	90	30.41	SRT	0	1.18
## 49	Male	Meningioma	Supratentorial	70	0.97	SRT	1	1.41
## 49.1	Male	Meningioma	Supratentorial	70	0.97	SRT	1	1.41
## 49.2	Male	Meningioma	Supratentorial	70	0.97	SRT	1	1.41
## 49.3	Male	Meningioma	Supratentorial	70	0.97	SRT	1	1.41
## 80	Female	Meningioma	Infratentorial	70	2.94	SRS	0	1.54
## 41	Female	Meningioma	Supratentorial	80	1.57	SRT	0	2.03

## 47	Male	HG glioma	Supratentorial	80	29.27	SRT	0	5.15
## 35	Female	Meningioma	Supratentorial	90	7.30	SRT	0	5.51
## 38	Male	HG glioma	Supratentorial	90	9.95	SRT	1	6.23
## 5	Male	HG glioma	Supratentorial	90	5.06	SRT	1	6.30
## 5.1	Male	HG glioma	Supratentorial	90	5.06	SRT	1	6.30
## 85	Male	HG glioma	Supratentorial	80	19.81	SRT	1	6.39
## 16	Female	HG glioma	Supratentorial	70	11.38	SRS	1	7.05
## 36	Female	Other	Supratentorial	70	14.26	SRT	0	7.18
## 36.1	Female	Other	Supratentorial	70	14.26	SRT	0	7.18
## 2	Male	HG glioma	Supratentorial	90	19.35	SRT	1	8.98
## 2.1	Male	HG glioma	Supratentorial	90	19.35	SRT	1	8.98
## 2.2	Male	HG glioma	Supratentorial	90	19.35	SRT	1	8.98
## 53	Male	HG glioma	Supratentorial	90	2.50	SRT	1	9.77
## 69	Female	Meningioma	Supratentorial	80	6.93	SRS	0	10.10
## 69.1	Female	Meningioma	Supratentorial	80	6.93	SRS	0	10.10
## 69.2	Female	Meningioma	Supratentorial	80	6.93	SRS	0	10.10
## 33	Female	Other	Infratentorial	70	13.45	SRT	1	10.82
## 74	Female	HG glioma	Supratentorial	80	0.19	SRT	1	11.02
## 10	Male	HG glioma	Supratentorial	100	2.53	SRT	0	11.48
## 50	Female	LG glioma	Supratentorial	80	0.19	SRS	0	11.51
## 27	Male	Other	Supratentorial	60	34.64	SRT	1	11.57
## 88	Male	Other	Infratentorial	80	0.11	SRT	0	13.90
## 22	Male	HG glioma	Supratentorial	70	14.44	SRT	1	14.00
## 43	Female	Meningioma	Supratentorial	70	6.70	SRT	0	14.56
## 43.1	Female	Meningioma	Supratentorial	70	6.70	SRT	0	14.56
## 37	Female	Meningioma	Supratentorial	80	6.60	SRT	0	14.75
## 42	Male	HG glioma	Supratentorial	90	0.28	SRT	1	16.43
## 42.1	Male	HG glioma	Supratentorial	90	0.28	SRT	1	16.43
## 55	Female	HG glioma	Supratentorial	80	0.63	SRT	1	16.92
## 55.1	Female	HG glioma	Supratentorial	80	0.63	SRT	1	16.92
## 60	Female	Meningioma	Supratentorial	90	2.47	SRT	0	17.57
## 71	Male	Meningioma	Supratentorial	70	8.45	SRT	0	19.41
## 58	Male	HG glioma	Supratentorial	80	3.75	SRT	1	19.90
## 58.1	Male	HG glioma	Supratentorial	80	3.75	SRT	1	19.90
## 58.2	Male	HG glioma	Supratentorial	80	3.75	SRT	1	19.90
## 58.3	Male	HG glioma	Supratentorial	80	3.75	SRT	1	19.90
## 73	Male	LG glioma	Supratentorial	90	2.64	SRT	0	20.13
## 73.1	Male	LG glioma	Supratentorial	90	2.64	SRT	0	20.13
## 84	Male	HG glioma	Supratentorial	80	0.16	SRT	1	20.69
## 84.1	Male	HG glioma	Supratentorial	80	0.16	SRT	1	20.69
## 68	Male	Other	Infratentorial	70	0.01	SRS	0	23.67
## 21	Female	Meningioma	Supratentorial	80	2.82	SRS	0	24.39
## 21.1	Female	Meningioma	Supratentorial	80	2.82	SRS	0	24.39
## 26	Male	LG glioma	Supratentorial	90	4.23	SRT	1	25.02
## 26.1	Male	LG glioma	Supratentorial	90	4.23	SRT	1	25.02
## 3	Female	Meningioma	Infratentorial	70	7.95	SRS	0	26.46
## 3.1	Female	Meningioma	Infratentorial	70	7.95	SRS	0	26.46
## 3.2	Female	Meningioma	Infratentorial	70	7.95	SRS	0	26.46
## 20	Female	Meningioma	Supratentorial	90	2.50	SRT	0	30.16
## 72	Male	Meningioma	Supratentorial	80	20.93	SRT	1	31.15
## 86	Male	Meningioma	Supratentorial	90	2.50	SRT	0	32.82
## 13	Female	Meningioma	Infratentorial	90	0.63	SRS	0	32.98
## 25	Female	Meningioma	Supratentorial	70	6.48	SRT	1	33.41
## 57	Male	HG glioma	Infratentorial	80	0.22	SRS	0	33.67

```
## 9      Female Meningioma Supratentorial 70 12.16 SRT 0 34.66
## 29      Male Meningioma Supratentorial 60 3.81 SRT 0 36.10
## 29.1    Male Meningioma Supratentorial 60 3.81 SRT 0 36.10
## 29.2    Male Meningioma Supratentorial 60 3.81 SRT 0 36.10
## 54      Male Meningioma Infratentorial 80 24.41 SRT 0 39.54
## 15      Female Meningioma Supratentorial 60 9.18 SRT 0 41.44
## 87      Male Meningioma Supratentorial 90 2.02 SRS 0 42.07
## 8       Male LG glioma Supratentorial 80 12.37 SRT 0 42.10
## 32      Male Meningioma Supratentorial 90 2.56 SRS 0 44.39
## 40      Female Meningioma Supratentorial 90 2.54 SRT 0 45.74
## 40.1    Female Meningioma Supratentorial 90 2.54 SRT 0 45.74
## 81      Female HG glioma Supratentorial 80 15.45 SRT 1 46.16
## 82      Female Other Supratentorial 90 1.82 SRT 0 47.11
## 4       Female LG glioma Supratentorial 80 7.61 SRT 1 47.80
## 24      Female Meningioma Infratentorial 100 2.13 SRS 1 51.02
## 24.1    Female Meningioma Infratentorial 100 2.13 SRS 1 51.02
## 24.2    Female Meningioma Infratentorial 100 2.13 SRS 1 51.02
## 6       Female Meningioma Supratentorial 80 4.82 SRS 0 52.75
## 6.1     Female Meningioma Supratentorial 80 4.82 SRS 0 52.75
## 34      Male Other Infratentorial 80 6.81 SRS 0 57.11
## 76      Female Meningioma Supratentorial 80 31.74 SRT 0 57.25
## 30      Female Meningioma Supratentorial 90 4.72 SRS 0 65.02
## 65      Male LG glioma Supratentorial 80 9.58 SRT 0 78.75
## 18      Male HG glioma Supratentorial 90 10.80 SRT 0 82.56
## 18.1    Male HG glioma Supratentorial 90 10.80 SRT 0 82.56
## 18.2    Male HG glioma Supratentorial 90 10.80 SRT 0 82.56
## 18.3    Male HG glioma Supratentorial 90 10.80 SRT 0 82.56
```

```
merge(x = braincancer_by_time, y = bootstrap_df_sample,
      by = 'time', all.x = T)
```

```
##      time  survive
## 1  0.07      NA
## 2  1.18 1.0000000
## 3  1.41 0.9540230
## 4  1.54 0.9540230
## 5  2.03 0.9540230
## 6  3.38      NA
## 7  4.16      NA
## 8  4.56      NA
## 9  5.15 0.9540230
## 10 5.51 0.9540230
## 11 6.10      NA
## 12 6.23 0.9419467
## 13 6.30 0.9177943
## 14 6.39 0.9057180
## 15 6.82      NA
## 16 6.92      NA
## 17 7.05 0.8936418
## 18 7.18 0.8936418
## 19 7.25      NA
## 20 8.98 0.8564067
## 21 9.77 0.8439950
## 22 10.10 0.8439950
```

```
## 23 10.49      NA
## 24 10.82 0.8310105
## 25 11.02 0.8180259
## 26 11.48 0.8180259
## 27 11.51 0.8180259
## 28 11.57 0.8046157
## 29 13.90 0.8046157
## 30 14.00 0.7909781
## 31 14.56 0.7909781
## 32 14.62      NA
## 33 14.75 0.7909781
## 34 16.43 0.7622153
## 35 16.92 0.7334524
## 36 17.57 0.7334524
## 37 18.95      NA
## 38 19.41 0.7334524
## 39 19.74      NA
## 40 19.90 0.6735788
## 41 20.13 0.6735788
## 42 20.69 0.6422495
## 43 22.03      NA
## 44 23.67 0.6422495
## 45 24.39 0.6422495
## 46 25.02 0.6084469
## 47 26.46 0.6084469
## 48 26.85      NA
## 49 29.70      NA
## 50 30.16 0.6084469
## 51 31.15 0.5894330
## 52 31.25      NA
## 53 31.67      NA
## 54 32.82 0.5894330
## 55 32.82 0.5894330
## 56 32.98 0.5894330
## 57 33.41 0.5691077
## 58 33.67 0.5691077
## 59 34.26      NA
## 60 34.66 0.5691077
## 61 35.93      NA
## 62 36.10 0.5691077
## 63 39.25      NA
## 64 39.54 0.5691077
## 65 39.54 0.5691077
## 66 41.44 0.5691077
## 67 42.07 0.5691077
## 68 42.10 0.5691077
## 69 44.39 0.5691077
## 70 45.74 0.5691077
## 71 46.16 0.5335385
## 72 47.11 0.5335385
## 73 47.80 0.4954286
## 74 49.05      NA
## 75 50.85      NA
## 76 51.02 0.3810989
```

```
## 77 52.23      NA
## 78 52.75 0.3810989
## 79 54.43      NA
## 80 55.80      NA
## 81 57.11 0.3810989
## 82 57.25 0.3810989
## 83 57.64      NA
## 84 65.02 0.3810989
## 85 67.38      NA
## 86 73.74      NA
## 87 78.75 0.3810989
## 88 82.56 0.3810989
```

```
set.seed(42)
```

```
B = 200
```

```
results = braincancer_by_time
```

```
for (i in 1:B){
```

```
  #print(i)
```

```
  bootstrap_index = sample(nrow(BrainCancer), size = 88, replace = T)
```

```
  a_bootstrap_sample = BrainCancer[bootstrap_index, ]
```

```
  a_bootstrap_km_curve = survfit(Surv(time, status) ~ 1,
                                data = a_bootstrap_sample,
                                conf.int = 0.68)
```

```
  a_bootstrap_df = data.frame(time = a_bootstrap_km_curve$time,
                              survive = a_bootstrap_km_curve$surv)
```

```
  results = merge(x = results, y = a_bootstrap_df,
                  by = 'time', all.x = T)
```

```
}
```

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y' are duplicated in the result
```

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y' are duplicated in the result
```

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y', 'survive.x', 'survive.y' are duplicated
## in the result
```

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y', 'survive.x', 'survive.y' are duplicated
## in the result
```

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y' are duplicated in the result
```

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y' are duplicated in the result
```

[illegible]





[illegible]







```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',  
## 'survive.y' are duplicated in the result
```

[illegible][illegible][illegible][illegible]

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x'
```



```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y' are duplicated
## in the result
```

[illegible][illegible][illegible][illegible]





```
## 'survive.y' are duplicated in the result
```

```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y' are duplicated
## in the result
```

[illegible]

[illegible][illegible]

```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',  
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y' are duplicated  
## in the result
```

[illegible][illegible][illegible]

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
```



[illegible][illegible]





[illegible][illegible]



[illegible][illegible]

[illegible][illegible]



```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y' are duplicated in the result
```

[illegible]



```
## 'survive.y' are duplicated in the result
```

[illegible][illegible]





```
## 'survive.y' are duplicated in the result
```

[illegible]

[illegible][illegible]

```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y' are duplicated
## in the result
```

[illegible][illegible]

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
```

[illegible][illegible][illegible]

[illegible]

[illegible][illegible]

```
## 'survive.y', 'survive.x', 'survive.y' are duplicated in the result
```

[illegible]









[illegible][illegible]















[illegible][illegible]

[illegible][illegible]





[illegible][illegible]

```
## in the result
```

[illegible]



[illegible][illegible]

[illegible][illegible]

[illegible]

```
## Warning in merge.data.frame(x = results, y = a_bootstrap_df, by = "time", :
## column names 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
```

[illegible][illegible]

```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y' are duplicated in the result
```

[illegible]



[illegible][illegible]

```
## 'survive.y' are duplicated in the result
```

[illegible]







```
## in the result
```

[illegible][illegible]



[illegible][illegible]

```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y' are duplicated
## in the result
```

[illegible]

[illegible]

[illegible]







[illegible][illegible]

[illegible][illegible]



[illegible]

```
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y', 'survive.x', 'survive.y', 'survive.x', 'survive.y', 'survive.x',
## 'survive.y' are duplicated in the result
```

```
# View(results)
```

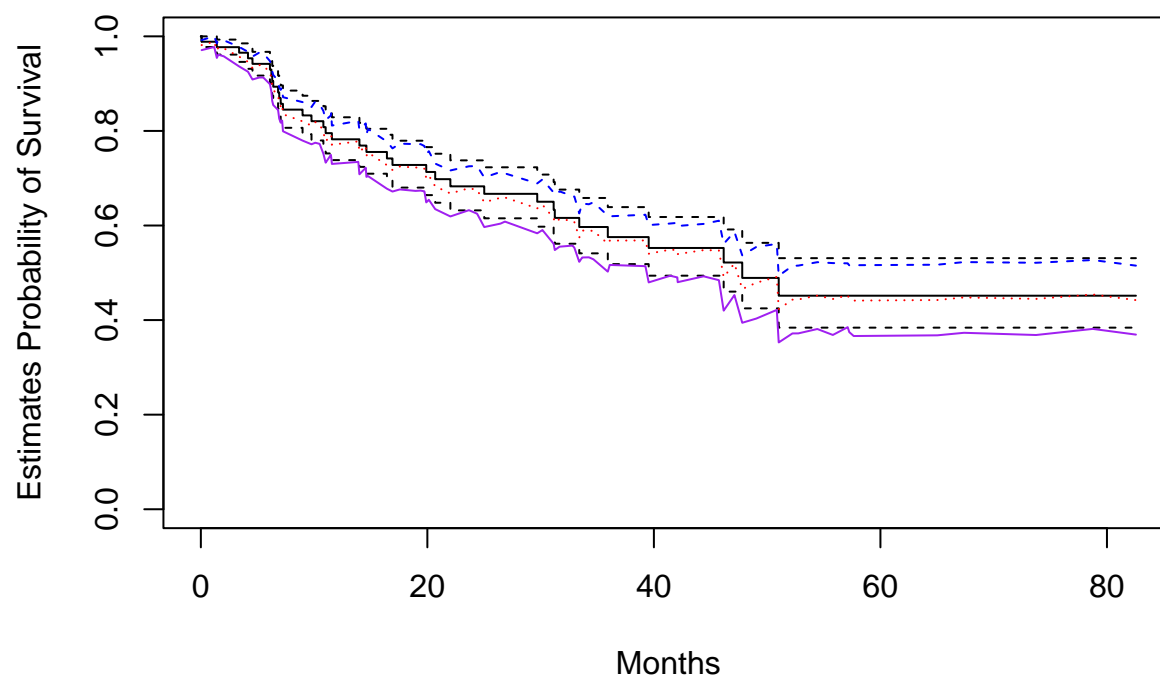
```
for (i in 1:nrow(BrainCancer)){

  a_row = na.omit(as.numeric(results[i, 2:201]))
  mean_prob = mean(a_row)
  std = sd(a_row)
  upper_bound = mean_prob + 1 * std
  lower_bound = mean_prob - 1 * std
  results[i, 'mean'] = mean_prob
  results[i, 'upper_bound'] = upper_bound
  results[i, 'lower_bound'] = lower_bound

}
```

```
View(results)
```

```
plot(km_68_curve, xlab = 'Months', ylab = 'Estimates Probability of Survival')
lines(x = results$time, y = results$mean, col = 'red', lty = 9)
lines(x = results$time, y = results$upper_bound, col = 'blue', lty = 8)
lines(x = results$time, y = results$lower_bound, col = 'purple', lty = 7)
```



c

```
cox_model = coxph(Surv(time, status) ~ ., data = BrainCancer)
cox_model
```

```
## Call:
## coxph(formula = Surv(time, status) ~ ., data = BrainCancer)
##
##              coef exp(coef) se(coef)      z      p
## sexMale         0.18375    1.20171  0.36036  0.510 0.61012
## diagnosisLG glioma 0.91502    2.49683  0.63816  1.434 0.15161
## diagnosisHG glioma 2.15457    8.62414  0.45052  4.782 1.73e-06
## diagnosisOther     0.88570    2.42467  0.65787  1.346 0.17821
## locSupratentorial  0.44119    1.55456  0.70367  0.627 0.53066
## ki               -0.05496    0.94653  0.01831 -3.001 0.00269
## gtv               0.03429    1.03489  0.02233  1.536 0.12466
## stereoSRT         0.17778    1.19456  0.60158  0.296 0.76760
##
## Likelihood ratio test=41.37  on 8 df, p=1.776e-06
## n= 87, number of events= 35
## (1 observation deleted due to missingness)
```

There are two statistically significant predictors: HG glioma and ki.



Patients diagnosed with HG glioma have a probability of event occurring (die) 8.6 times higher than patients who are not diagnosed with HG glioma. (lower probability of survival)

Each additional “ki” leads to an 0.95-fold increase in the risk of being died.

d

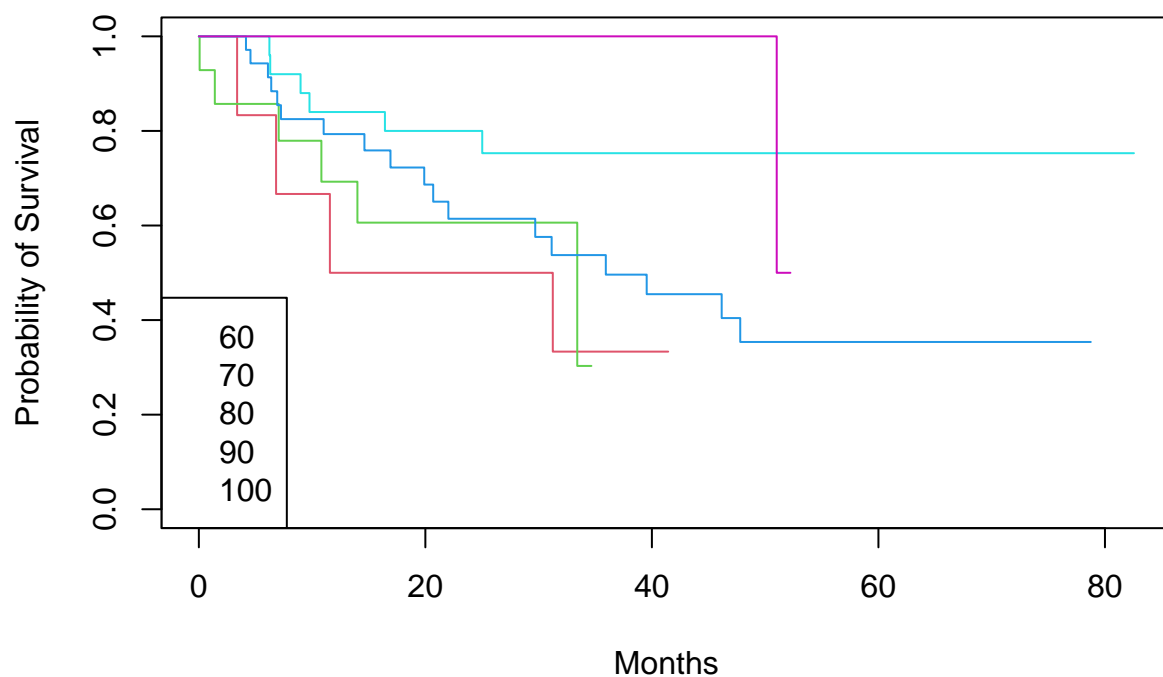
```
braincancer = BrainCancer
braincancer$ki_group = (cut(BrainCancer$ki, breaks = c(0, 60, 70, 80, 90, 100),
                           labels = c('60', '70', '80', '90', '100')))
table(BrainCancer$ki)
```

```
##
##  40  60  70  80  90 100
##   1   5  14  36  27   5
```

```
table(braincancer$ki_group)
```

```
##
##  60  70  80  90 100
##   6  14  36  27   5
```

```
ki_model = survfit(Surv(time, status) ~ ki_group, data = braincancer)
plot(ki_model, xlab = 'Months', ylab = 'Probability of Survival',
     col = c(2, 3, 4, 5, 6))
legend('bottomleft', legend = levels(braincancer$ki_group),
     col = c(2, 3, 4, 5, 6))
```



## Exercise 11

a

```
Observation = c(26.5, 37.2, 57.3, 90.8, 20.2, 89.8)
Censoring_Indicator = c(1, 1, 1, 0, 0, 0)
Covariate = c(0.1, 11, -0.3, 2.8, 1.8, 0.4)

data = data.frame(observation = Observation,
                  censoring_indicator = Censoring_Indicator,
                  covariate = Covariate)

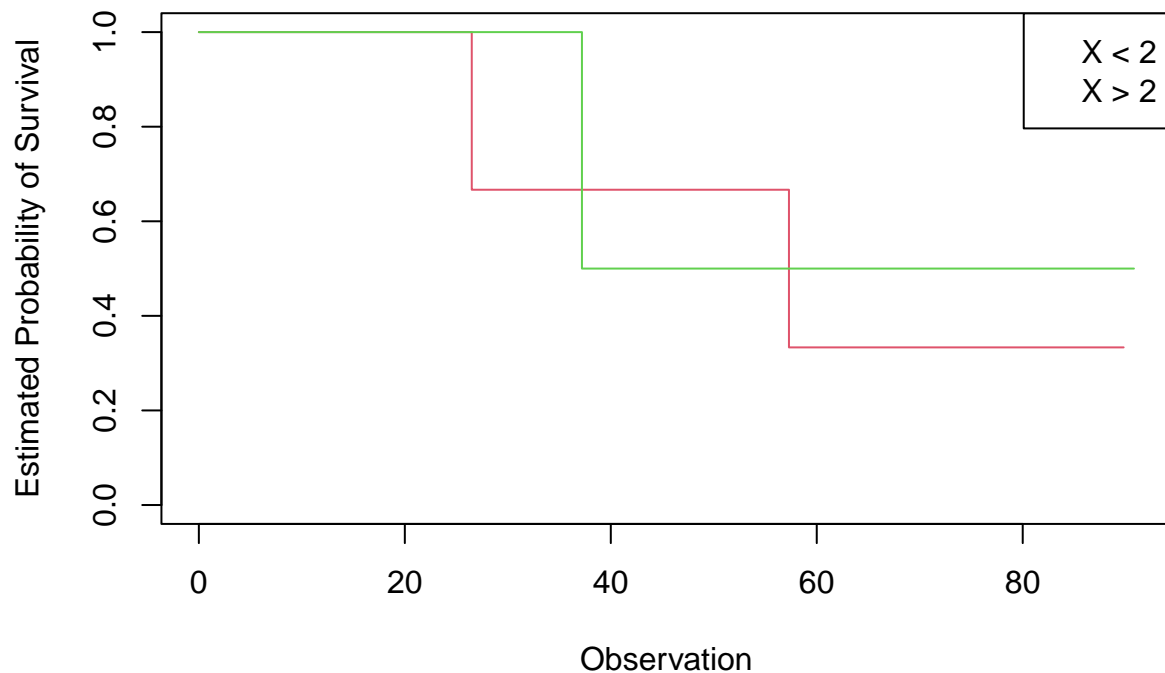
data
```

```
##   observation censoring_indicator covariate
## 1      26.5              1         0.1
## 2      37.2              1        11.0
## 3      57.3              1        -0.3
## 4      90.8              0         2.8
## 5      20.2              0         1.8
## 6      89.8              0         0.4
```

```

data$covariate_group = cut(data$covariate, breaks = c(-Inf, 2, Inf),
                           labels = c('X < 2', 'X > 2'))
covariate_curve = survfit(Surv(observation, censoring_indicator) ~ covariate_group,
                         data = data)
plot(covariate_curve, xlab = 'Observation',
     ylab = 'Estimated Probability of Survival',
     col = c(2, 3))
legend('topright', legend = c('X < 2', 'X > 2'), col = c(2, 3))

```



b

```

cox_model = coxph(Surv(observation, censoring_indicator) ~ covariate_group,
                  data = data)
cox_model

```

```

## Call:
## coxph(formula = Surv(observation, censoring_indicator) ~ covariate_group,
##       data = data)
##
##               coef exp(coef) se(coef)      z      p
## covariate_groupX > 2 -0.3401    0.7117  1.2359 -0.275 0.783
##
## Likelihood ratio test=0.08  on 1 df, p=0.7797

```

```
## n= 6, number of events= 3
```

Objects in group number 2 (where X is greater than or equal 2) have a risk of event occurring 0.7117-fold higher than Objects in group number 1.

Since p-value is greater than 0.05 (approximately 0.78), the data provides convincing evidence that the true coefficient is non-zero. (There is a difference between 2 groups).

**c**

```
log_rank_test = survdiff(Surv(observation, censoring_indicator) ~ covariate_group,  
                          data = data)  
log_rank_test
```

```
## Call:  
## survdiff(formula = Surv(observation, censoring_indicator) ~ covariate_group,  
##          data = data)  
##  
##              N Observed Expected (O-E)^2/E (O-E)^2/V  
## covariate_group=X < 2 4          2      1.77    0.0308    0.0764  
## covariate_group=X > 2 2          1      1.23    0.0441    0.0764  
##  
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.8
```

The p-value results from Cox model and log-rank test are identical. The result above was rounded.