Chapter 4. Data Plots Peter smith

```r
knitr::include_graphics("4_1.PNG")
```

4.1. Estimate the probability of survival beyond a full 3-year term in office for the 'Months in office for Australian Prime Ministers' data of Example 4.1. Place appropriate error bounds on your estimate.

Required data

```r
## Months in office for Australian Prime Ministers

ministers <- c(seq(1:29))
months <- c(33,8,5,12,41,8,11,39,16,14,89,81,28,88,1,29,3,46,1,54,194,24,2,39,22,36,89,106,52)
```

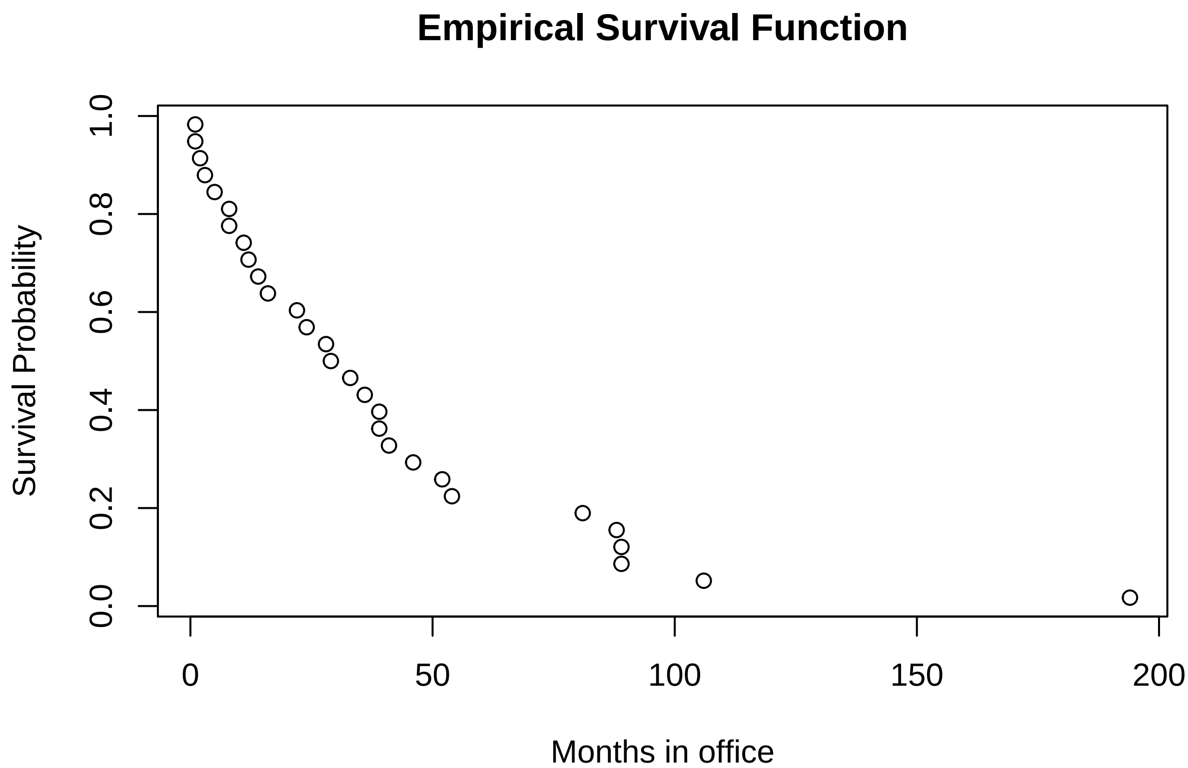
Plotting position calculations

```r
## Plotting position

p <- (ministers - 0.5)/length(ministers)
head(p)
```

```
## [1] 0.01724138 0.05172414 0.08620690 0.12068966 0.15517241 0.18965517
```

Data Plot

```r
## Empirical survivor plot

months <- sort(months)
data <- as.data.frame(cbind(months,1-p))
plot(months,1-p, main="Empirical Survival Function", xlab = "Months in office", ylab = "Survival Probab
```



Estimate probablity of survival beyond full 3 years

Knowing that the estimate probability of the survival fucntion is giving by

$$S_n\left(y\right) = \frac{number\ of\ observations\ > y}{n} = \frac{1}{n}\sum_{i=1}^{n}\ I\left(y,\infty\right)\left(Y_i\right)$$

then:

$$S_{29}\left(17\right) = \frac{1}{29}\sum_{i=1}^{n}\ I\left(12,\infty\right)\left(Y_i\right) = \frac{12}{29}\ =\ 0.4137$$

Confidence interval

And the approximate confidence interval based in two standard error is:

$$S_{29}\left(17\right) \pm 2\sqrt{\frac{S_{29}\left(12\right)\left(1-S_{29}\left(12\right)\right)}{n}}$$

then

$$0.4137 + 2\sqrt{\frac{0.4137\left(1-0.4137\right)}{29}} = 0.5966$$

and

$$0.4137 - 2\sqrt{\frac{0.4137\,(1-0.4137)}{29}} = 0.2308$$

```
knitr::include_graphics("4_2.PNG")
```

4.2. Construct the empirical distribution function for 'Survival (days)' for the patients observed to die from 'heart failure' in the 1980 Stanford Heart Transplant Data from Crowley and Hu (1977). (The complete data set will be discussed in later chapters and is listed in Example 5.2.)

Give a point and interval estimate for the probability of survival beyond 1000 days.

Required data

```
## Stanford Heart Transplat Data

## Reading data
heart <- read.csv("heart_data.csv", header=T,sep=";")
head(heart)
```

```
##   Days Cens  Age   T5
## 1   15    1 54.3 1.11
## 2    3    1 40.4 1.66
## 3  624    1 51.0 1.32
## 4   46    1 42.5 0.61
## 5  127    1 48.0 0.36
## 6   64    1 54.6 1.89
```

Cleaning the censored data

```
heart_cens <- subset(heart, Cens == 1)
attach(heart_cens)
heart_cens
```

```
##    Days Cens  Age     T5
## 1    15    1 54.3   1.11
## 2     3    1 40.4   1.66
## 3   624    1 51.0   1.32
## 4    46    1 42.5   0.61
## 5   127    1 48.0   0.36
## 6    64    1 54.6   1.89
## 7  1350    1 54.1   0.87
## 8   280    1 49.5   1.12
## 9    23    1 56.9   2.05
## 10   10    1 55.3   2.76
## 11 1024    1 43.4   1.13
```

```
## 12   39    1 42.8    1.38
## 13  730    1 58.4    0.96
## 14  136    1 52.0    1.62
## 16    1    1 54.2    0.47
## 17  836    1 45.0    1.58
## 18   60    1 64.5    0.69
## 21   54    1 49.0    2.09
## 22   47    1 61.5    0.87
## 23   51    1 50.5 9999.00
## 26   44    1 36.2    0.00
## 27  994    1 48.6    0.81
## 28   51    1 47.2    1.38
## 30  897    1 46.1 9999.00
## 31  253    1 48.8    1.08
## 32  147    1 47.5 9999.00
## 33   51    1 52.5    1.51
## 35  322    1 48.1    1.82
## 37   65    1 49.1    0.66
## 39  551    1 48.9    0.12
## 40   66    1 51.3    1.12
## 41  228    1 19.7    1.02
## 42   65    1 45.2    1.68
## 44   25    1 53.0    1.68
## 47   63    1 56.4    2.16
## 48   12    1 29.2    0.61
## 51   29    1 54.0    1.08
## 54   48    1 53.4    3.05
## 55  297    1 42.8    0.60
## 57   50    1 46.4    2.25
## 59   68    1 51.4    1.33
## 60   26    1 52.5    0.82
## 63  161    1 43.8    1.20
## 64   14    1 40.3 9999.00
## 69    1    1 41.5    0.87
```

Plotting position calculations

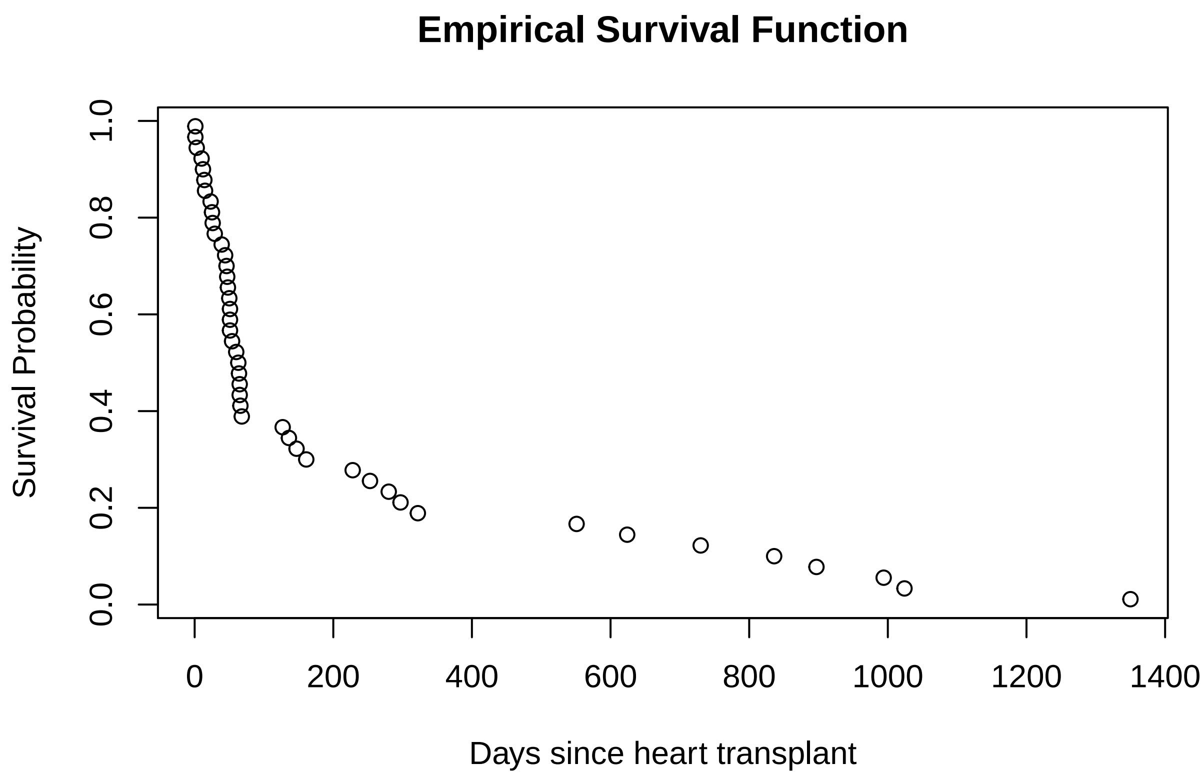
```r
## Plotting position

patients <- seq(1:length(Days))
p <- (patients - 0.5)/length(patients)
head(p)
```

```
## [1] 0.01111111 0.03333333 0.05555556 0.07777778 0.10000000 0.12222222
```

Data Plot

```r
## Empirical survivor plot

Days <- sort(Days)
```

```r
plot(Days,1-p, main="Empirical Survival Function", xlab = "
Days since heart transplant", ylab = "Survival Probability")
```



Estimate probablity of survival beyond full 3 years

```r
## Looking on the data greater than 1000

sum(Days > 1000)
```

```
## [1] 2
```

Then by the definition 1.3:

$$S_{45}\left(43\right)=\frac{1}{45}\sum\,_{i=1}^{45}I\left(2,\infty\,\right)\left(Y_i\right)=\frac{2}{45}=0.0444$$

Confidence interval

The approximate confidence interval based in two standard error is:

$$0.4444+2\sqrt{\frac{0.4444\left(1-0.4444\right)}{45}}=0.5925$$

and

$$0.4444 - 2\sqrt{\frac{0.4444\,(1-0.4444)}{45}} = 0.2962$$

```r
knitr::include_graphics("4_3.PNG")
```

4.3. Construct a hazard plot for 'Survival (days)' for the patients observed to die from 'heart failure' in the 1980 Stanford Heart Transplant Data from Crowley and Hu (1977). (The complete data set will be discussed in later chapters and is listed in Example 5.2.)

The defitnion 4.6 says that the hazard plot is given by:

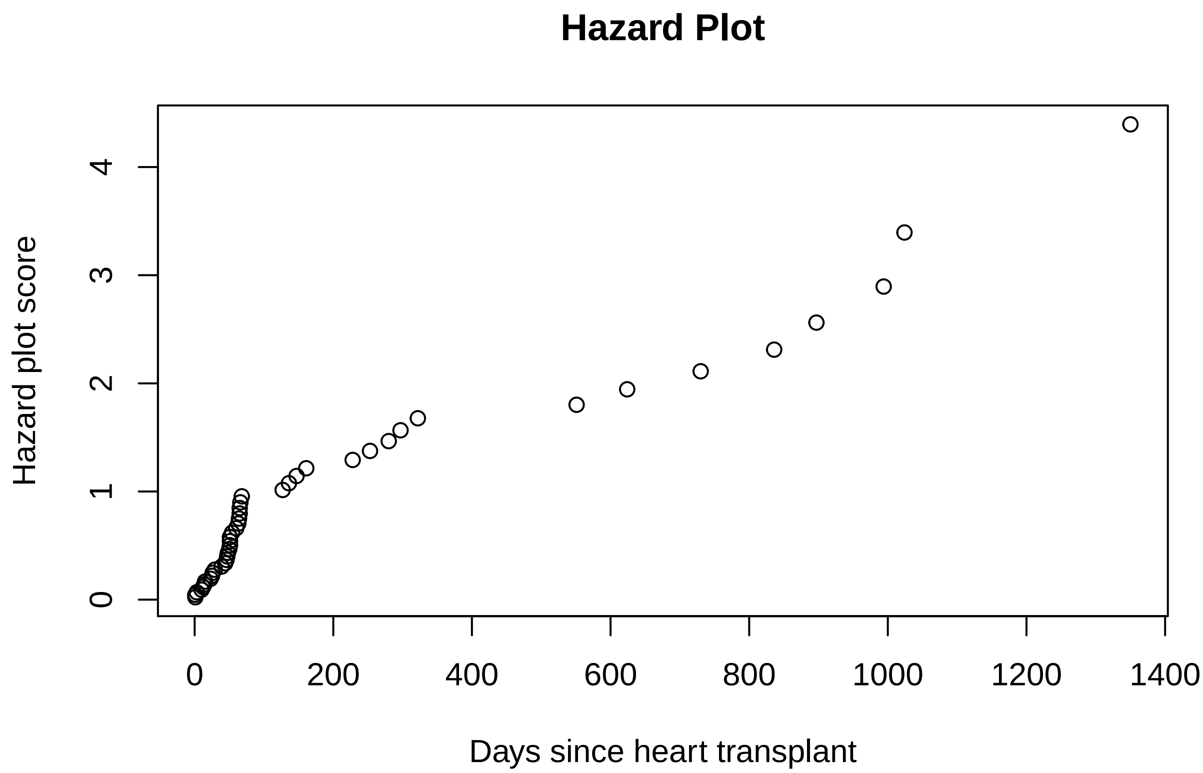
$$\left(\alpha_i, y_{(j)}\right)$$

Where

$$\alpha_j = \sum{}_{i=1}^{j}\left(\frac{1}{n-i+1}\right)$$

```r
alpha <-c()
sumatory <- c()

for(j in 1:length(Days)){

  for(i in 1:j){

    sumatory[i]<- (1/(length(Days)-i+1))

  }
  alpha[j]<- sum(sumatory)
  sumatory <- 0
}
```

```r
plot(Days,alpha, main="Hazard Plot", xlab = "
Days since heart transplant", ylab = "Hazard plot score")
```




```r
knitr::include_graphics("4_3_1.PNG")
```

Give an approximate point and interval estimate of the cumulative hazard at 1000 days.

knowing that

$$H_{n(y)} = -ln\left(S_n\left(y\right)\right)$$

We have that:

$$H_{45}\left(43\right) = -ln\left(0.0444\right) = 3.1145$$

```r
knitr::include_graphics("4_3_2.PNG")
```

Use a probability plot to assess whether survival time after transplant follows a lognormal distribution.