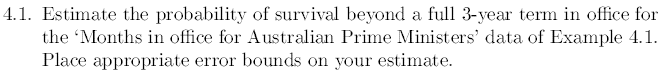
Chapter 4. Data Plots Peter smith

### Required libraries

library(matlib)

## Warning: package 'matlib' was built under R version 4.0.4

knitr::include\_graphics("4\_1.PNG")



### Required data

To facilitate the calculations, we will plot the empirical survival function, first creating the vectors that store the data.

## 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)

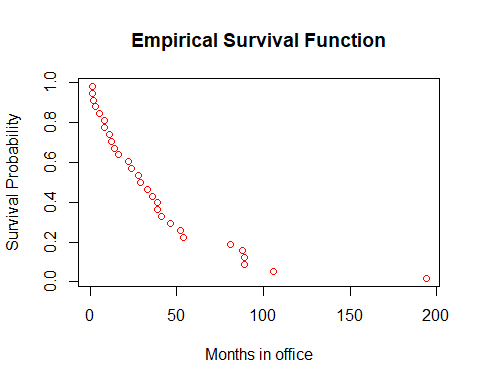
### Plotting position calculations

## 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

## 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 Probability", col ="red")



### Estimate probablity of survival beyond full 3 years

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

then:

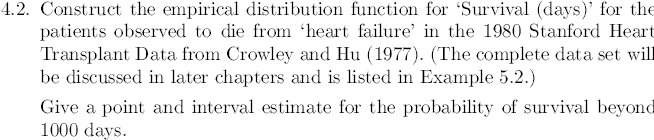
### Confidence interval

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

then

and

knitr::include\_graphics("4\_2.PNG")



### Required data

## Stanford Heart Transplat Data  
  
## Reading data  
heart <- read.csv("heart\_data.csv", header=T,sep=";")  
attach(heart)  
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

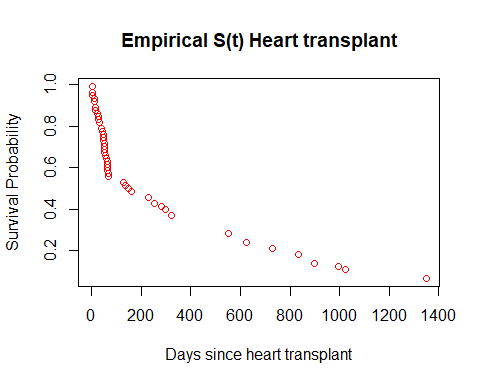
### Plotting position calculations

## Plotting position  
  
patients <- seq(1:length(Days))  
p <- (patients - 0.5)/length(patients)  
head(p)

## [1] 0.007246377 0.021739130 0.036231884 0.050724638 0.065217391 0.079710145

### Data Plot

## Empirical survivor plot  
  
i <- order(heart$Days)  
heart <- heart[i,]  
p <- 1-p  
  
Days <- heart$Days  
Cens <- heart$Cens  
  
newHeart <- as.data.frame(cbind(Days,Cens,p))  
newHeart <- subset(newHeart, Cens==1)  
plot(newHeart$Days,newHeart$p, main="Empirical S(t) Heart transplant", xlab = "  
Days since heart transplant", ylab = "Survival Probability", col="red")



### Estimate probablity of survival beyond 1000 days

## Looking on the data greater than 1000  
  
sum(heart$Days > 1000)

## [1] 8

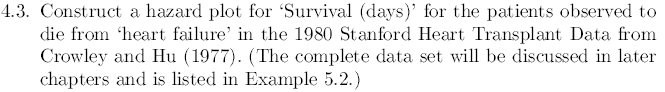
Then by the definition 1.3:

#### Confidence interval

The approximate confidence interval based in two standard error is:

and

knitr::include\_graphics("4\_3.PNG")

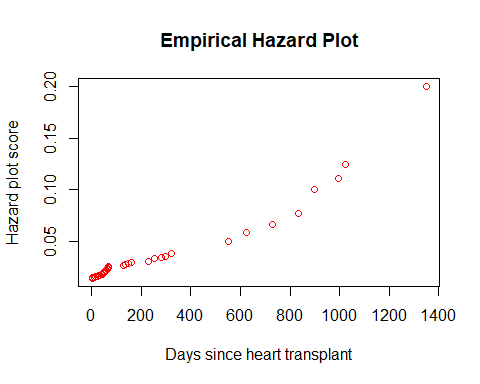
 ### Calculation of the empirical and accumulative Hazard

k <- seq(69,1)  
Emp.h <- c()  
Days <- heart$Days   
Cens <- heart$Cens  
  
### The following cycles calculate all the data for the empirical risk, for ease of calculation the censored data are filled with zeros.  
  
  
for(i in 1:69){  
   
 if(Cens[i] == 0){  
   
 Emp.h[i]= 0  
  
   
 }else{  
   
 Emp.h[i] = 1/k[i]  
  
 }  
   
}  
  
cum.H <- cumsum(Emp.h)  
  
for(i in 1:69){  
   
 if(Cens[i] == 0){  
   
 cum.H[i]= 0  
  
 }  
}  
  
  
harzard.score<- as.data.frame(cbind(Days,Cens, k, Emp.h, cum.H))  
  
head(harzard.score)

## Days Cens k Emp.h cum.H  
## 1 1 1 69 0.01449275 0.01449275  
## 2 1 0 68 0.00000000 0.00000000  
## 3 1 1 67 0.01492537 0.02941813  
## 4 3 1 66 0.01515152 0.04456964  
## 5 10 1 65 0.01538462 0.05995426  
## 6 12 1 64 0.01562500 0.07557926

### Building the Empirical harzard plot

### Building the graph  
  
hz.plotting <- subset(harzard.score, Cens == 1)  
plot(hz.plotting$Days,hz.plotting$Emp.h, main="Empirical Hazard Plot", xlab = "Days since heart transplant", ylab = "Hazard plot score", col="red")



knitr::include\_graphics("4\_3\_1.PNG")



### Estimated accrued risk at 1,000 days, the closest point is 994 days.  
  
hz.plotting$cum.H[hz.plotting$Days == 994]

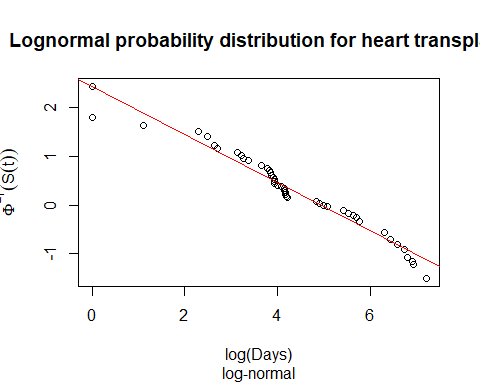
## [1] 1.291767

Based on the calculations in the table above, we can say that the accumulated empirical hazard at 1000 days is 1.2918.

knitr::include\_graphics("4\_3\_2.PNG")

 ### Calculations of lognormal probability distribution

inv\_phi <- qnorm(newHeart$p)  
result.lm <- lm(inv\_phi~log(newHeart$Days))  
  
plot(log(newHeart$Days),inv\_phi, main = "Lognormal probability distribution for heart transplant",xlab="log(Days)", sub="log-normal",ylab=expression(Phi^-1 \* (S(t))))  
abline(result.lm, col="red")



result.lm

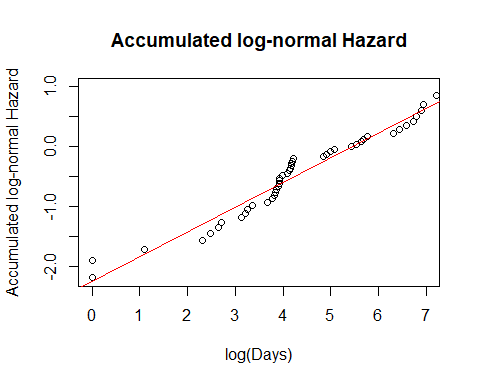
##   
## Call:  
## lm(formula = inv\_phi ~ log(newHeart$Days))  
##   
## Coefficients:  
## (Intercept) log(newHeart$Days)   
## 2.4423 -0.4948

Using the graph as empirical evidence, it can be corroborated that the data are generated to fit a lognormal distribution( 2.4423 , -0.4948 ).

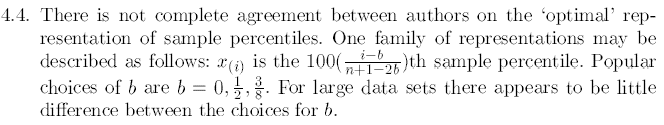
### Accumulated hazard calculations

Similarly, we can calculate the cumulative hazard for a log normal distribution as follows:

Hexp <- 1-exp(-hz.plotting$cum.H)  
Hphi <- qnorm(Hexp)  
  
LogNomrh.lm <- lm(Hphi~log(hz.plotting$Days))  
  
plot(log(hz.plotting$Days),Hphi, main="Accumulated log−normal Hazard", ylab = "Accumulated log−normal Hazard", xlab="log(Days)", xlim=c(0,7), ylim=c(-2.2,1) )  
abline(LogNomrh.lm, col="red")



knitr::include\_graphics("4\_4.PNG")

 In our case the author has decided to use b = 0.5, as can be seen below: