

## HW#1 Survival analysis II, Spring 2015, Due 3/3(Tue)

### Name:

1. Read and summarize Sec. 1.3 (p.7-)
2. a) Read and summarize Sec. 1.5 (p.9-)
  - b) Use mathematical formulas to explain what is done in the codes
  - c) Calculate the unconditional mean lifetime by simulations

### Answer by Pan Chi-Hung:

#### Section 1.3:

This section is talk about how to input the data form text file to R language. In data analysis, we should use data to perform our method. Sometimes the data is not available form some package, so we should construct it. The book gives a process.

First, the data ( Table 1.3 ) should be set up in a text file, say catheter1.dat.

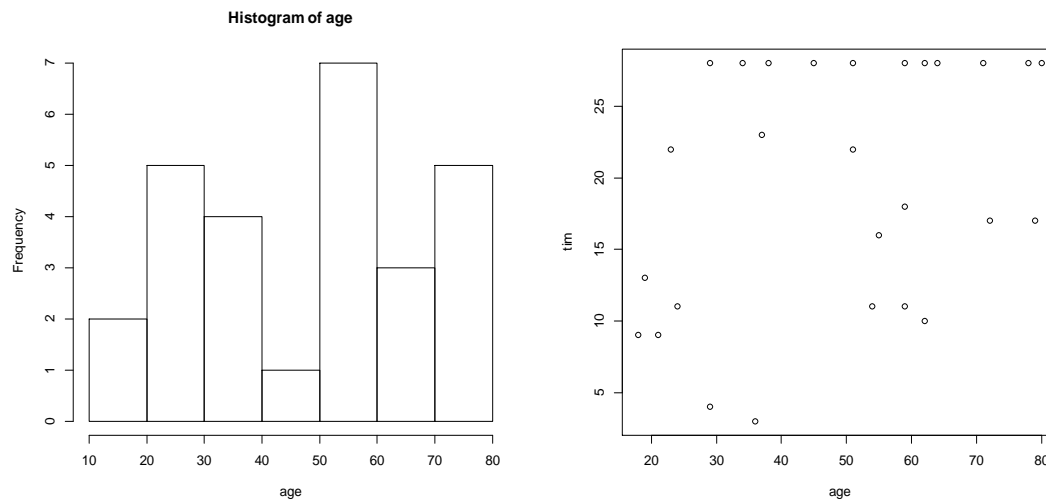
Second, loaded the data into R. Following is the code.

```
##input the data and named dmx
dmx = read.table("catheter1.dat", header=T)
attach(dmx)
dmx
##some function
summary(dmx)
hist(age)
plot(age,tim)
```

The function attach will make each column has a name( Title ). Also these are many functions in R. For example, mean(), var(), summary(), sort(), hist(),...,plot().

### Output:

```
> summary(dmx)
      id          age          sex          tim
Min.   : 1.0    Min.   :18.00   Min.   :1.000   Min.   : 3.00
1st Qu.: 7.5    1st Qu.:31.50   1st Qu.:1.000   1st Qu.:11.00
Median :14.0    Median :51.00   Median :2.000   Median :22.00
Mean   :14.0    Mean   :48.48   Mean   :1.519   Mean   :19.41
3rd Qu.:20.5    3rd Qu.:62.00   3rd Qu.:2.000   3rd Qu.:28.00
Max.   :27.0    Max.   :80.00   Max.   :2.000   Max.   :28.00
      cns
Min.   :0.0000
1st Qu.:0.0000
Median :0.0000
Mean   :0.4815
3rd Qu.:1.0000
Max.   :1.0000
```



### Section1.5:

Simulation is very useful to check our statistical inference is correct or not. This section is talk about how to generate the data what we want. Now we want to generate some right-censored survival times whose depends on some recorded factors. Suppose that  $T_i$ , the breakdown time of the  $i$  machine, has an exponential distribution with mean a given function of  $x_1$ , a measure of the intensity of usage, and  $x_2$  a measure of the frequency and quality of maintenance. This book give the R code is as follows:

```
n=25; b0=1.5; b1=1.2; b2=-2.5
x1=rep(0,n); x2=x1;tim=x1
for( i in 1:n ){ x1[i]=runif(1,min=0,max=1); x2[i]=runif(1);
                lam=exp(b0+b1*x1[i]+b2*x2[i]);
                ti=rexp(1,rate=lam);
                tim[i]=min(ti,10);}

for( i in 1:n ){ v1=c(x1[i],x2[i],tim[i]);
                v2=format(v1,width=8,digits=2);
                cat("\n",v2);}
```

n: sample size

The code means that:

$T | \mathbf{x}$  is from the distribution  $1 - \exp\{-\lambda(\mathbf{x})t\}$ ,

where  $\mathbf{x} = (x_1, x_2)$  and  $\lambda(\mathbf{x}) = \exp(b_0 + b_1x_1 + b_2x_2)$ , where  $x_1$  and  $x_2$  are

form  $\text{uniform}(0, 1)$ . This sentence  $\text{tim}[i]=\min(\text{ti},10)$  in the code means that only

$T | \mathbf{x} < 10$  we can get the data otherwise set  $T | \mathbf{x} = 10$ . Finally, these are 25 samples

under right-censored at 10. The last for loop is just combine the  $\mathbf{x}$  and the censored

sample.

```

n=50000; b0=1.5; b1=1.2; b2=-2.5
x1=rep(0,n); x2=x1;tim=x1
for( i in 1:n ){ x1[i]=runif(1,min=0,max=1); x2[i]=runif(1);
                lam=exp(b0+b1*x1[i]+b2*x2[i]);
                ti=rexp(1,rate=lam);
                tim[i]=min(ti,10);}

lamda = exp(b0+b1*x1+b2*x2)
mean(tim)
mean(1/lamda)

```

Under censored assumption:

$$E(T) \approx \sum_i T_i / n = 0.573274 \quad (\text{calculate it by simulation, sample size is 50000})$$

If there is no censored assumption:

$$\begin{aligned}
E(T) &= E\{E(T|\mathbf{x})\} = E\{1/\lambda(x)\} \\
&= E\{1/\exp(b_0 + b_1x_1 + b_2x_2)\} \approx \sum_i \lambda(\mathbf{x}_i)/n = 0.5764391
\end{aligned}$$

**Output:**

```

> mean(tim)
[1] 0.573274
> mean(1/lamda)
[1] 0.5764391

```