PSTAT 175: Week 3

Shaocong Ma

October 17, 2018

Today

- `quantile()` function
- Confident Interval
  - Use `survfit()` function
  - Greenwood formula
- `survfit()` for different groups
- Lab B: 3(d). Hypothesis test

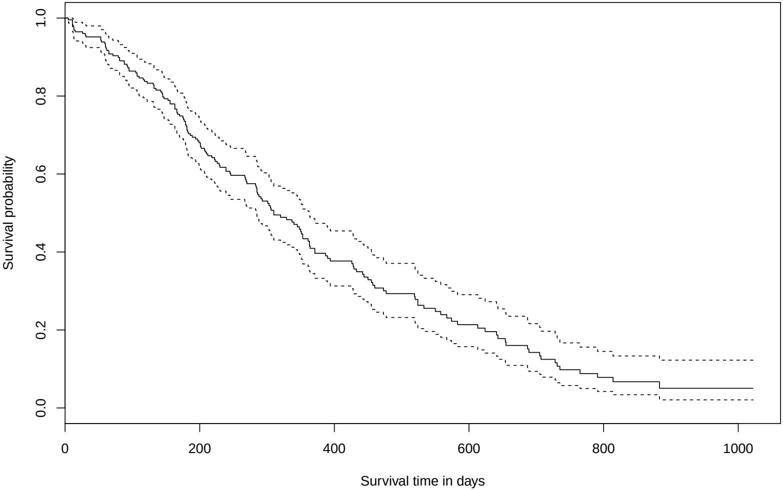
```
heroin = read.table("Heroin.txt")
heroin.time = heroin$Time
heroin.cns = heroin$Status
heroin.surv <- Surv(heroin.time, heroin.cns)
heroin$Group <- ifelse(heroin$Time <= 365, 1, 0)
#factor(heroin$Group, levels=c(1,2))
#survdiff(heroin.surv ~ heroin$Group, rho=0)
#km = survfit(heroin.surv ~ heroin$Group)

sum(heroin$Group)
```

Review: `survfit()` function

Create a survival object and plot KM estimator with 95% CI.

```
lungkm = survfit(Surv(time,status)~1,data=lung)
plot(lungkm,xlab="Survival time in days",
     ylab="Survival probability" )
```

Review: `survfit()` function



quantile() function

Compute the 10th quantile (the first time the survival function is below .9):

```
min(lungkm$time[lungkm$surv < .9])
```

```
## [1] 79
```

`quantile()` function

Use `quantile()` funtion in `R`

```
quantile(lungkm,probs = .1, conf.int = FALSE)
```

```
## 10
## 79
```

`quantile()` function

Use `quantile()` funtion in `R`

```
quantile(lungkm,probs =c(.1,.2,.75), conf.int = FALSE)
```

```
##  10  20  75
##  79 145 550
```

Confident Interval

Confident interval for estimated survival probability:

```
summary(lungkm)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, data = l
##
##  time n.risk n.event survival std.err lower 95% CI upper
##     5    228       1   0.9956 0.00438       0.9871
##    11    227       3   0.9825 0.00869       0.9656
##    12    224       1   0.9781 0.00970       0.9592
##    13    223       2   0.9693 0.01142       0.9472
##    15    221       1   0.9649 0.01219       0.9413
##    26    220       1   0.9605 0.01290       0.9356
##    30    219       1   0.9561 0.01356       0.9299
##    31    218       1   0.9518 0.01419       0.9243
##    53    217       2   0.9430 0.01536       0.9134
##    54    215       1   0.9386 0.01590       0.9079
##    59    214       1   0.9342 0.01642       0.9026
```

Confident Interval

```r
s = summary(lungkm)
names(s)
```

```
##  [1] "n"               "time"            "n.risk"          "n.
##  [5] "n.censor"        "surv"            "type"            "st
##  [9] "lower"           "upper"           "conf.type"       "co
## [13] "call"            "table"           "rmean.endtime"
```

```r
#s$lower
#s$upper
```

Greenwood formula

95% CI for $\log S(t)$:

$$\log \hat{S}(t) \pm 1.96 \hat{S}(t) \sqrt{\sum_{j=1}^{k} \frac{m_j}{n_j(n_j - m_j)}}$$

95% CI for $S(t)$:

$$\hat{S}(t) \times \exp\left[\pm 1.96 \hat{S}(t) \sqrt{\sum_{j=1}^{k} \frac{m_j}{n_j(n_j - m_j)}} \right]$$

Compute it by hand

```
mj = lungkm$n.event
nj = lungkm$n.risk
```

```
Vj = mj/nj/(nj-mj)
cVj = cumsum(Vj)
```

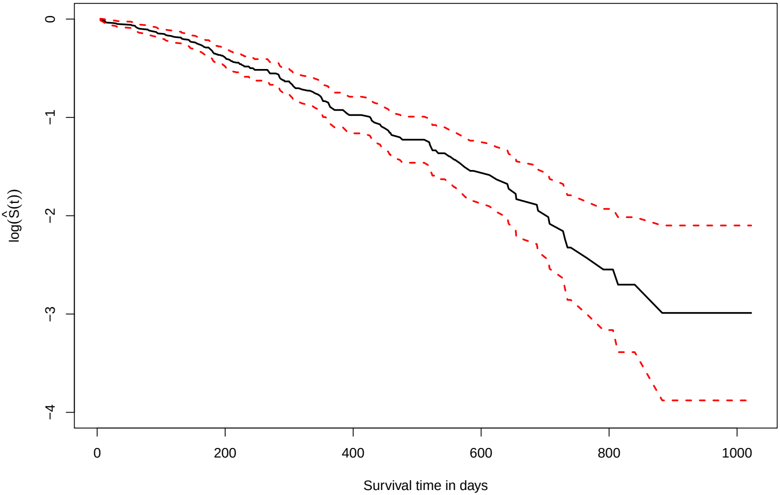
Compute it by hand

Reminder: 95% CI for $\log S(t)$:

$$\log \hat{S}(t) \pm 1.96\hat{S}(t)\sqrt{\sum_{j=1}^{k} \frac{m_j}{n_j(n_j - m_j)}}$$

```
lowerCI = log(lungkm$surv) - 1.96*sqrt(cVj)
upperCI = log(lungkm$surv) + 1.96*sqrt(cVj)
```

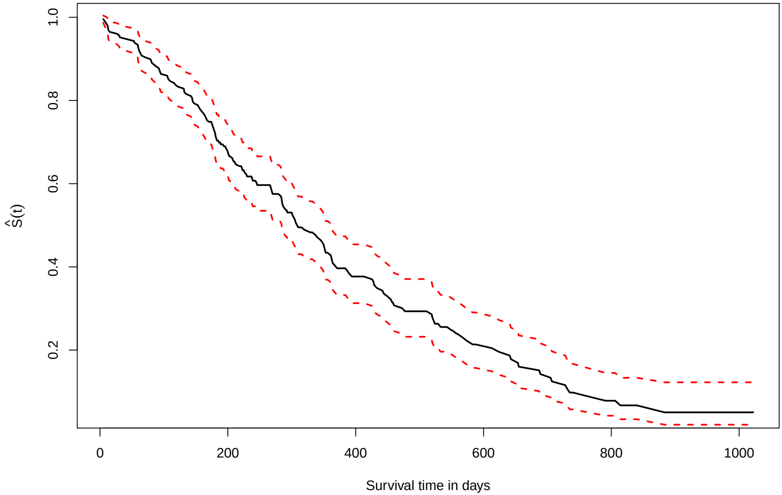
Plot it



Transform it into CI for S(t)

```r
par(mar=c(5,5,4,2))
plot(lungkm$time,lungkm$surv,lwd=2,type="l",
xlab="Survival time in days",ylab=expression(hat(S)(t)))
lines(lungkm$time,exp(lowerCI),lty=2,col=2,lwd=2)
lines(lungkm$time,exp(upperCI),lty=2,col=2,lwd=2)
```

Transform it into CI for S(t)



`survfit()` for different groups

Divide data into to part. Treat them separetely

```
g1 = lung[lung$sex==1,]
g2 = lung[lung$sex==2,]
kmg1 = survfit(Surv(time,status)~1,data=g1)
kmg2 = survfit(Surv(time,status)~1,data=g2)
```

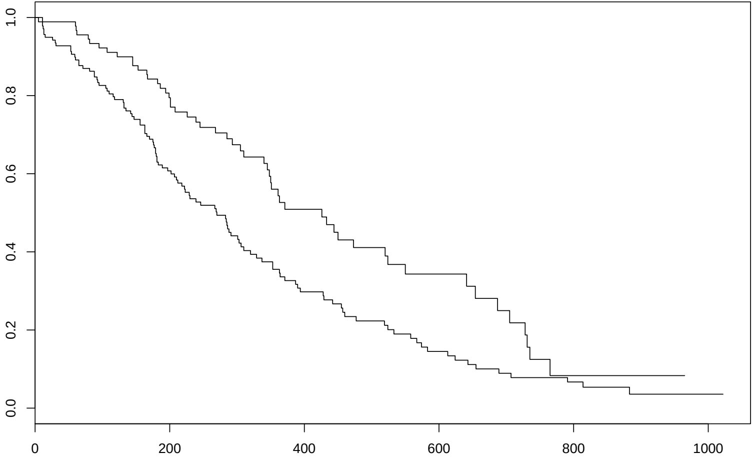
`surfit()` function

Change 1 to `sex`:

```
km = survfit(Surv(time,status)~sex,data=lung)
```

Plot

```r
plot(km)
```



Last part: 3(d)

The file `heroin.Rdt` contains data from a study on in-patient methadone treatment clinics in Australia. The columns are labeled `Status` and `Time` which gives the number of days each subject spent in the clinics. Censored observations were generally subjects who were still in the clinics at the end of the study period.

(d)New recommendations for clinic administration are that, in order to save money, at least 50% of the patients should be discharged within one year. Is there significant evidence that most patients from this study population are in the clinics for more than one year? Perform a hypothesis test using the relevant statistic and an approximation to its standard error. Should we use a one-sided or a two-sided alternative? Compute an approximate P-value for the test.