

PSTAT 175: Week 5

Shaocong Ma

October 31, 2018

Today

- ▶ Review: K-M estimate
- ▶ Review: Compare two groups
 - ▶ `survdif()` function
 - ▶ Plot K-M estimates for two groups
- ▶ `coxph()` function
 - ▶ Cox proportional hazards model
 - ▶ Likelihood ratio test

veteran dataset

In today's section, we use veteran data set in survival package from Veterans' Administration Lung Cancer study.

```
head(veteran)
```

##	trt	celltype	time	status	karno	diagtime	age	prior
## 1	1	squamous	72	1	60	7	69	0
## 2	1	squamous	411	1	70	5	64	10
## 3	1	squamous	228	1	60	3	38	0
## 4	1	squamous	126	1	60	9	63	10
## 5	1	squamous	118	1	70	11	65	10
## 6	1	squamous	10	1	20	5	49	0

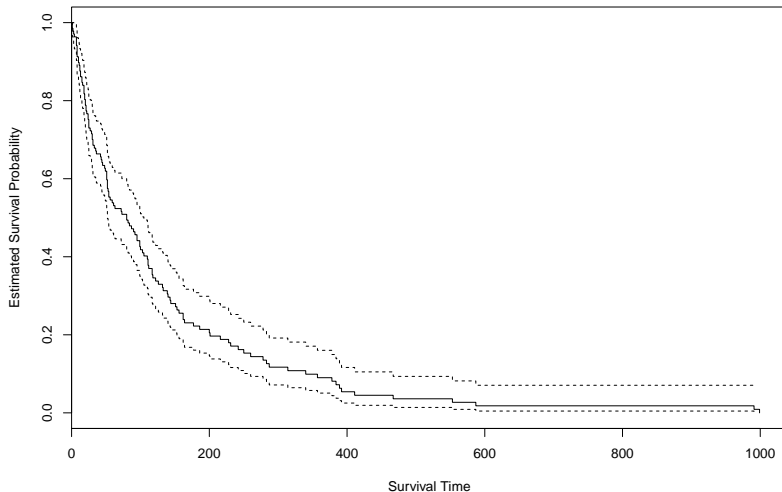
veteran dataset

- ▶ trt: 1=standard 2=test
- ▶ celltype: 1=squamous, 2=smallcell, 3=adeno, 4=large
- ▶ time: survival time
- ▶ status: censoring status
- ▶ karno: Karnofsky performance score (100=good)
- ▶ diagtime: months from diagnosis to randomisation
- ▶ age: in years
- ▶ prior: prior therapy 0=no, 10=yes

See [R: Veterans' Administration Lung Cancer study](#) for more details.

Review: K-M estimate

We only consider time and status. Plot the Kaplan–Meier estimate of the survivor function.



Review: K-M estimate

Code:

```
veteran.km = survfit(Surv(time,status)~1, data=veteran)
plot(veteran.km,xlab="Survival Time",
      ylab="Estimated Survival Probability" )
```

Review: Compare two groups

Goal: To study the effect of treatment. Divide all observations into two groups based on trt.

```
survdif(Surv(time,status)~trt, data=veteran)
```

```
## Call:
```

```
## survdiff(formula = Surv(time, status) ~ trt, data = vete
```

```
##
```

```
##           N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## trt=1 69      64      64.5   0.00388   0.00823
```

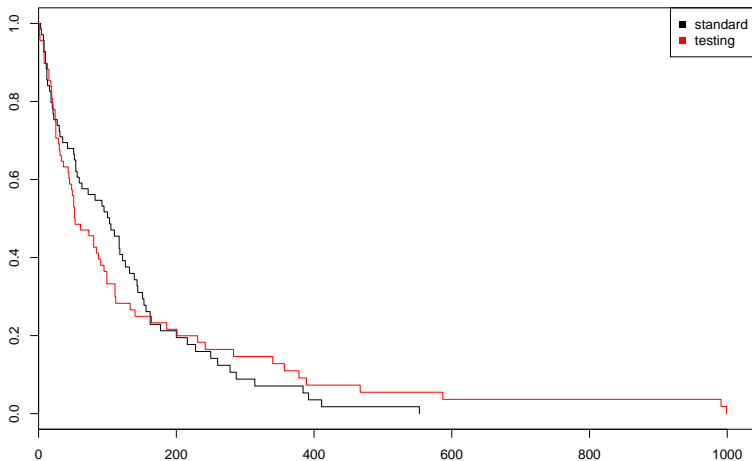
```
## trt=2 68      64      63.5   0.00394   0.00823
```

```
##
```

```
##  Chisq= 0   on 1 degrees of freedom, p= 0.9
```

Review: Compare two groups

The p-value is 0.9. It means there is no significant difference between the standard group and the test group. Now, plot the K-M estimators for both groups:



Review: Compare two groups

- ▶ When $t > 200$, the estimated survival probability in testing group is always larger than that in standard group.
- ▶ Log-rank test claims that two groups are same.

coxph() function

Use coxph() function to compare two groups. Do not consider covariates.

```
coxph(Surv(time,status)~trt,data=veteran)
```

```
## Call:
```

```
## coxph(formula = Surv(time, status) ~ trt, data = veteran
```

```
##
```

```
##           coef exp(coef) se(coef)      z      p
```

```
## trt 0.0177      1.0179   0.1807 0.1 0.92
```

```
##
```

```
## Likelihood ratio test=0.01  on 1 df, p=0.9
```

```
## n= 137, number of events= 128
```

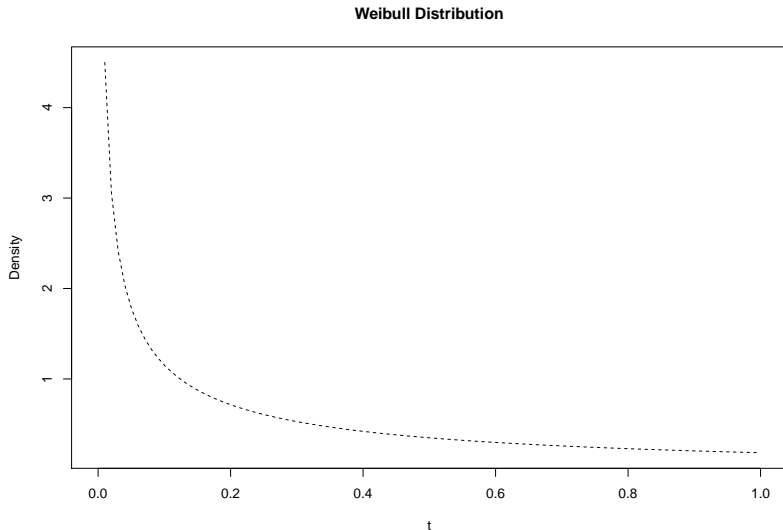
coxph() function

Use coxph() function to compare two groups. Consider covarites age, celltype, karno, and diagtime to control for differences in the groups.

```
#including `trt`  
cox = coxph(Surv(time,status)~age+celltype+karno+diagtime  
            +trt, data=veteran)  
  
#not including `trt`  
cox2 = coxph(Surv(time,status)~age+celltype+karno+diagtime  
            ,data=veteran)
```

Cox proportional hazards model

Reminder: Hazard function can be considered as the risk of dying at time t . For example, $h(t)$ for leukemia patients has Weibull distribution.



Cox proportional hazards model

Semiparametric model for hazard function:

$$h(t, X) = h_0(t)e^{\sum_{i=1}^p \beta_i X_i}.$$

- ▶ $h_0(t)$ is called *the baseline hazard function*.
- ▶ Proportional hazards assumption: $h_0(t)$ only relies on t .
- ▶ Time-independence.

Likelihood ratio test

```
cox$loglik
```

```
## [1] -505.4491 -474.4443
```

The first one is for null model where there is no covariate. The second one is what we need.

Likelihood ratio test

Reminder:

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L_{\theta}(x)}{\sup_{\theta \in \Theta} L_{\theta}(x)}$$

By monotonicity of log function:

$$\log \lambda(x) = \sup_{\theta \in \Theta_0} l_{\theta}(x) - \sup_{\theta \in \Theta} l_{\theta}(x)$$

Recall the asymptotic property:

$$-2 \log \lambda \xrightarrow{D} \chi_1^2$$

Likelihood ratio test

- Compute the Likelihood ratio

```
lrt = 2*(cox$loglik[2]-cox2$loglik[2])  
lrt
```

```
## [1] 2.071351
```

- It has approximated chi-square distribution with degree 1

```
pchisq(lrt,df=1,lower.tail=FALSE)
```

```
## [1] 0.1500885
```


Comment: coxph() function

call:

```
coxph(formula = Surv(time, status) ~ age + celltype + karno +  
      diagtime + trt, data = veteran)
```

n= 137, number of events= 128

	coef	exp(coef)	se(coef)	z	Pr(> z)	
age	-0.008706	0.991332	0.009309	-0.935	0.34971	
celltypesmallcell	0.851206	2.342471	0.273011	3.118	0.00182	**
celltypeadeno	1.183667	3.266330	0.297896	3.973	7.08e-05	***
celltypelarge	0.401001	1.493318	0.282665	1.419	0.15600	
karno	-0.032586	0.967940	0.005447	-5.982	2.21e-09	***
diagtime	0.001339	1.001340	0.008066	0.166	0.86814	
trt	0.298380	1.347674	0.207503	1.438	0.15045	

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
age	0.9913	1.0087	0.9734	1.0096
celltypesmallcell	2.3425	0.4269	1.3718	4.0000
celltypeadeno	3.2663	0.3062	1.8218	5.8564
celltypelarge	1.4933	0.6696	0.8581	2.5987
karno	0.9679	1.0331	0.9577	0.9783
diagtime	1.0013	0.9987	0.9856	1.0173
trt	1.3477	0.7420	0.8973	2.0240

Concordance= 0.738 (se = 0.03)

Rsquare= 0.364 (max possible= 0.999)

Likelihood ratio test= 62.01 on 7 df, p=6e-11

wald test = 62.41 on 7 df, p=5e-11

score (logrank) test = 66.74 on 7 df, p=7e-12

Figure 1: summary(cox)

Construct CI for parameters

```
confint(cox,level=.95)
```

##	2.5 %	97.5 %
## age	-0.02695186	0.009540464
## celltypesmallcell	0.31611466	1.386298277
## celltypeadeno	0.59980068	1.767533126
## celltypelarge	-0.15301235	0.955013877
## karno	-0.04326240	-0.021908783
## diagtime	-0.01446981	0.017148119
## trt	-0.10831833	0.705079044