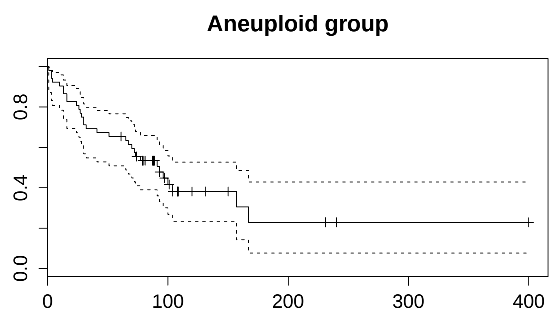
APTS - Survival Analysis
Lab Session 1 - Solutions

Ingrid Van Keilegom

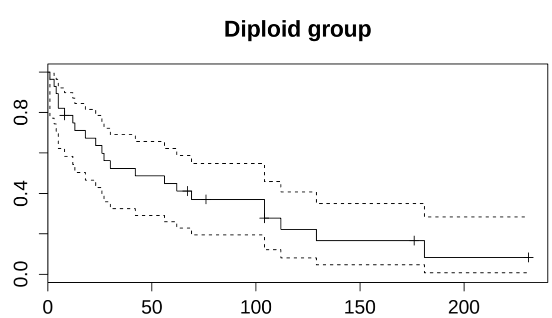
August 23, 2017

1.
```
> install.packages("survival")
> library("survival")
> install.packages("KMsurv")
> library("KMsurv")
> data(tongue)
> tongue

> ane=subset(tongue,type==1)
> fit_ane=survfit(Surv(time,delta)~1,data=ane,conf.type="log-log")
> plot(fit_ane)
> title(main="Aneuploid group")
```




```
> dip=subset(tongue,type==2)
> fit_dip=survfit(Surv(time,delta)~1,data=dip,conf.type="log-log")
> plot(fit_dip)
> title(main="Diploid group")
```



```
> fit_ane
Call: survfit(formula = Surv(time, delta) ~ 1, data = ane, conf.type = "log-log")

      n  events  median 0.95LCL 0.95UCL
     52      31      93      65     157
> fit_dip
Call: survfit(formula = Surv(time, delta) ~ 1, data = dip, conf.type = "log-log")

      n  events  median 0.95LCL 0.95UCL
     28      22      42      18     104
```

2.
```
> data(burn)
> burn

> survdiff(Surv(T3,D3)~Z1,data=burn)
Call:
survdiff(formula = Surv(T3, D3) ~ Z1, data = burn)

      N Observed Expected (O-E)^2/E (O-E)^2/V
Z1=0 70       28     21.4      2.07      3.79
Z1=1 84       20     26.6      1.66      3.79

 Chisq= 3.8  on 1 degrees of freedom, p= 0.0515

> attach(burn)
> burn$area[Z4<=29] = 1
> burn$area[Z4>=30 & Z4<=50] = 2
> burn$area[Z4>=51] = 3

> survdiff(Surv(T3,D3)~Z1+strata(area),data=burn)
Call:
survdiff(formula = Surv(T3, D3) ~ Z1 + strata(area), data = burn)

      N Observed Expected (O-E)^2/E (O-E)^2/V
Z1=0 70       28     21.6      1.87      3.61
Z1=1 84       20     26.4      1.53      3.61

 Chisq= 3.6  on 1 degrees of freedom, p= 0.0574
```

3. (a) (i) The number of distinct event times : 5

(ii) The ordered event times :

j	1	2	3	4	5
$y_{(j)}$	2	5	7	9	16

(iii) The size of the risk set at time $y_{(1)}$: $y_{(1)} = 2$ and $R_{(1)} = 10$ since all subjects are alive just before time $t = 2$

(b) Calculation by hand (without using `R`) :

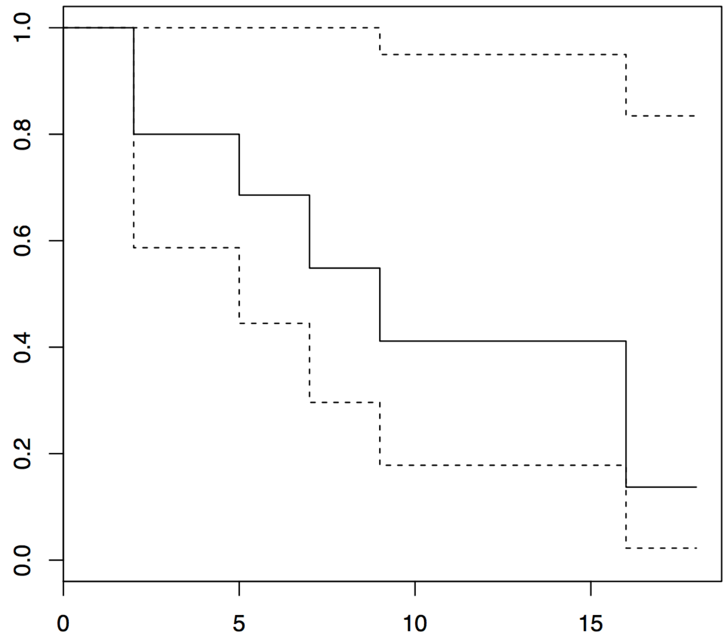
(i) Kaplan-Meier estimator : $\hat{S}_{KM}(t) = \prod_{j:y_{(j)} \leq t} \left(1 - \frac{d_{(j)}}{R_{(j)}}\right)$

(ii) Nelson-Aalen estimator : $\hat{S}_{NA}(t) = \prod_{j:y_{(j)} \leq t} \exp\left(-\frac{d_{(j)}}{R_{(j)}}\right)$

j	$y_{(j)}$	$d_{(j)}$	$R_{(j)}$	$1 - \frac{d_{(j)}}{R_{(j)}}$	$\hat{S}_{KM}(t)$	$\exp\left(-\frac{d_{(j)}}{R_{(j)}}\right)$	$\hat{S}_{NA}(t)$
1	2	2	10	0.800	0.800	0.819	0.819
2	5	1	7	0.857	0.686	0.867	0.71
3	7	1	5	0.800	0.549	0.819	0.581
4	9	1	4	0.750	0.411	0.779	0.523
5	16	2	3	0.333	0.137	0.513	0.232

(c) Graphical representation of the Kaplan-Meier estimator :

```
> install.package("survival")
> y <- c(3,5,7,2,18,16,2,9,16,5)
> d <- c(0,1,1,1,0,1,1,1,1,0)
> library(survival)
> KM = survfit(Surv(y,d)~1)
> plot(KM)
```



(d) Comparison of the results :

```
> summary(KM)

Call: survfit(formula = Surv(y, d) ~ 1)
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
 2      10        2 0.800    0.126        0.5868 1.000
 5       7        1 0.686    0.151        0.4447 1.000
 7       5        1 0.549    0.172        0.2963 1.000
 9       4        1 0.411    0.176        0.1782 0.950
 16      3        2 0.137    0.126        0.0225 0.834
```

4. (a) Log-rank test :

$$\begin{aligned} U &= \sum_{j=1}^{r} w(y_{(j)})\,(O_j - E_j) \\ &= \sum_{j=1}^{r} w(y_{(j)}) \left(d_{(j)1} - \frac{d_{(j)} R_{(j)1}}{R_{(j)}} \right) \end{aligned}$$

with $\frac{U}{\sqrt{Var(U)}} \sim N(0,1)$ and

$$Var(U) = \sum_{j=1}^{r} w^2(y_{(j)}) \frac{d_{(j)} \frac{R_{(j)1}}{R_{(j)}} \left(1 - \frac{R_{(j)1}}{R_{(j)}}\right) \left(R_{(j)} - d_{(j)}\right)}{R_{(j)} - 1}.$$

		group 1		group 2					$Var(U)$		
j	$y_{(j)}$	$d_{(j)1}$	$R_{(j)1}$	$d_{(j)2}$	$R_{(j)2}$	$d_{(j)}$	$R_{(j)}$	$E_{(j)}$	$N_{(j)}$	$D_{(j)}$	$N_{(j)}/D_{(j)}$
1	4.1	1	6	0	6	1	12	0.5	2.75	11	0.25
2	9.7	0	4	1	6	1	10	0.4	2.16	9	0.24
3	10	2	4	1	5	3	9	1.333	4.44	8	0.555
4	17.2	1	1	0	2	1	3	0.333	0.44	2	0.22
5	19.7	0	0	1	2	1	2	0	0.00	1	0.00
Total		4		3		7		2.566			1.265

Hence,

$$U^{obs} = \frac{U}{\sqrt{Var(U)}} = \frac{1.434}{\sqrt{1.265}} = 1.275$$

We reject H_0 if $|U^{obs}| > z_{1-\alpha/2} = 1.96$.
We have $|U^{obs}| = 1.275 < z_{1-\alpha/2} = 1.96$.
Hence, we do not reject H_0.
$P\text{-}value = 2 \times P(Z > 1.275) = 2 \times 0.101 = 0.202 > 0.05$

(b) Comparison of the results with the function `survdiff` :

```
> y = c(4.1,7.8,10,10,12.3,17.2,
+       9.7,10,11.1,13.1,19.7,24.1)
> d = c(1,0,1,1,0,1,
+       1,1,0,0,1,0)
> group = c( rep(1,6), rep(2,6) )
> cbind(y, d, group)

         y  d  group
[1,]   4.1  1      1
[2,]   7.8  0      1
[3,]  10.0  1      1
[4,]  10.0  1      1
[5,]  12.3  0      1
[6,]  17.2  1      1
[7,]   9.7  1      2
[8,]  10.0  1      2
[9,]  11.1  0      2
[10,] 13.1  0      2
[11,] 19.7  1      2
[12,] 24.1  0      2

> library(survival)
> survdiff(Surv(y,d)~group)
```

```
Call:
survdiff(formula = Surv(y, d) ~ group)

        N Observed Expected (O-E)^2/E (O-E)^2/V
group=1 6        4     2.57     0.800      1.62
group=2 6        3     4.43     0.463      1.62

 Chisq= 1.6  on 1 degrees of freedom, p= 0.203
```

The conclusion is the same.