Survival Models

Lecture II. Confidence intervals for Kaplan-Meier estimator. Comparision of survival functions

Variance of $\hat{S}(t)$. Greenwood's formula (1)

Besides point estimates also the variability of obtained estimates is of interest, thus we start constructing the variance for Kaplan-Meier estimator

Let us first recall the K-M formula

$$\hat{S}(t) = \prod_{i:y_{(i)} \leq t} \frac{n_i - d_i}{n_i}$$

Note that the number of deaths d_i can be considered as a realization of some rv D_i

Taking natural logarithms gives

$$\ln \hat{S}(t) = \sum_{i:y_{(i)} \leq t} \ln \left(\frac{n_i - d_i}{n_i} \right)$$

Variance of $\hat{S}(t)$. Greenwood's formula (2)

Now, because of independence, the variance of $\ln \hat{S}(t)$ is given by

$$Var(\ln \hat{S}(t)) = \sum_{i: y_{(i)} \leq t} Var\left(\ln\left(\frac{n_i - D_i}{n_i}\right)\right)$$

Notice also that by construction the rv D_i has binomial distribution, $D_i \sim Bin(n_i, \frac{d_i}{n_i})$, which implies that

$$Var\left(\frac{n_i - D_i}{n_i}\right) = \frac{n_i \frac{d_i}{n_i}(1 - \frac{d_i}{n_i})}{n_i^2} = \frac{d_i(n_i - d_i)}{n_i^3}$$

Variance of $\hat{S}(t)$. Greenwood's formula (3)

Recall also that the variance of a function of a rv can be approximated using delta-method as follows:

$$Var(g(X)) \approx VarX \cdot \left(\frac{d\, g(x)}{dx}\Big|_{x=EX} \right)^2$$

In our case this means

$$Var\left(\ln\left(\frac{n_i - D_i}{n_i}\right)\right) \approx Var\left(\frac{n_i - D_i}{n_i}\right) \cdot \left(\frac{n_i}{n_i - d_i}\right)^2$$

Combining this result with the last result from previous slide, we get

$$Var\left(\ln\left(\frac{n_i - D_i}{n_i}\right)\right) \approx \frac{d_i(n_i - d_i)}{n_i^3} \cdot \left(\frac{n_i}{n_i - d_i}\right)^2 = \frac{d_i}{n_i(n_i - d_i)}$$

Variance of $\hat{S}(t)$. Greenwood's formula (4)

Now, summing the last result over all i, we obtain the following formula:

$$Var(\ln \hat{S}(t)) = \sum_{i: y_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

On the other hand, applying the delta-method directly to $Var(\ln(\hat{S}(t))$, we get

$$Var(\ln \hat{S}(t)) \approx \frac{1}{[\hat{S}(t)]^2} Var(\hat{S}(t))$$

which implies

Greenwood's formula

$$\widehat{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{i: y_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

Variance estimates for cumulative hazard (1)

Using the K-M estimator we can write

$$\hat{H}(t) = -\ln \hat{S}(t) = -\ln \prod_{i:y_{(i)} \leq t} \frac{n_i - d_i}{n_i},$$

from where the estimate for variance is given by

$$\widehat{Var}\,\hat{H}(t) = \sum_{i:y_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

Variance estimates for cumulative hazard (2)

One can also construct the confidence intervals for cumulative hazard using the Nelson-Aalen estimator

$$\hat{H}(t) = \sum_{i:y_{(i)} \leq t} \frac{d_i}{n_i},$$

Then corresponding variance is given by

$$\widehat{Var}\,\hat{H}(t) = \sum_{i:y_{(i)} \leq t} \frac{d_i}{n_i^2},$$

Confidence intervals for $S(t)$. Plain intervals

Denote the standard error of $\hat{S}(t)$ by $s.e.(\hat{S}(t))$, i.e. $s.e.(\hat{S}(t)) = \sqrt{\widehat{Var}(\hat{S}(t))}$

The *plain* confidence intervals using the Greenwood's standard error are calculated as

$$\hat{S}(t) \pm z_{\frac{\alpha}{2}} s.e.(\hat{S}(t))$$

The problem with given confidence intervals is that there is no guarantee that the limits actually are inside $[0, 1]$.

In R programming language the estimate for survival function with corresponding confidence intervals can be calculated using `survfit` function. By default, however, intervals are not calculated using the plain approach. To obtain the plain intervals, one must specify `conf.type="plain"` option.

Confidence intervals for $S(t)$. Log-intervals

Another possibility is to calculate the confidence intervals through the cumulative hazard, which results in so-called *log*-intervals

$$\exp\left(\ln \hat{S}(t) \pm z_{\frac{\alpha}{2}} s.e.(\hat{H}(t))\right),$$

where $s.e.(\hat{H}(t)) = \sqrt{\widehat{Var}\,\hat{H}(t)}$ and $\widehat{Var}\,\hat{H}(t)$ is calculated using K-M approach

In R `survfit` function this is the default method for calculation of confidence intervals (i.e., by default `conf.type="log"`)

Obviously, the limits of such confidence intervals are always positive, but they can still exceed 1.

Confidence intervals for $S(t)$. Log-log-intervals

In order to ensure that the confidence limits are within $[0,1]$, so-called *log-log* intervals are proposed (Kalbfleisch & Prentice, 1980).

Let us first denote $W = \ln(-\ln \hat{S}(t))$, then applying the delta-method results in

$$\widehat{Var}(W) \approx \frac{1}{[\ln \hat{S}(t)]^2} Var(-\ln \hat{S}(t)) \approx \frac{1}{[\ln \hat{S}(t)]^2} \sum_{i:y_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

Now, the following confidence intervals are constructed

$$[\hat{S}(t)]^{\exp\{\mp z_{\frac{\alpha}{2}} s.e.(W)\}}$$

The confidence intervals constructed this way are guaranteed to have limits within $[0,1]$

In R `survfit` function the confidence intervals are calculated using this method if the option `conf.type="log-log"` is specified

Home assignment

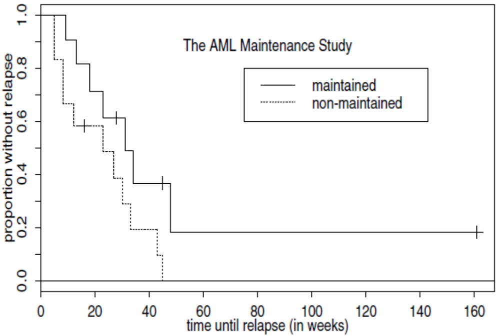
Home assignment 3.
Explain why the limits of the confidence interval for $S(t)$ constructed through the confidence interval for $W = \ln(-\ln \hat{S}(t))$ are of the form

$$[\hat{S}(t)]^{\exp\{\mp z_{\frac{\alpha}{2}}\, s.e.(W)\}}.$$

Show that both upper and lower limit of such confidence intervals are always within $[0, 1]$.

Comparison of survivor curves: Two-sample problem

Let us return to the AML example. The following figure compares the K-M survival curves for maintained and non-maintained group:



Question

How to make sure if the difference between two groups is signficant or not?

Comparison of two binomial populations

Setup:

- two populations (e.g., treatment/no treatment, method1/method2)
- each individual in either population can have one of two characteristics (e.g., dead/alive, failed/passed)

For example:

	Dead	Alive	
Population 1	a	b	n_1
Population 2	c	d	n_2
	m_1	m_2	n

Let us denote: $p_1 = P\{\text{Dead} \mid \text{Population1}\}$ and $p_2 = P\{\text{Dead} \mid \text{Population2}\}$.

Null hypothesis to test: $H_0 : p_1 = p_2$

Fisher's exact test (1)

Let us focus on the $(1,1)$ cell of the previous table. Consider its value a as a realization from random variable A.

Then, under H_0, given n_1, n_2, m_1, m_2 fixed quantities, the rv A has hypergeometric distribution:

$$P\{A = a\} = \frac{\binom{n_1}{a}\binom{n_2}{m_1 - a}}{\binom{n}{m_1}}$$

The test based on this exact distribution is called the Fisher's exact test.

Then, under H_0, the mean and variance are:

$$E_0 A = \frac{n_1 m_1}{n}$$

$$Var_0(A) = \frac{n_1 n_2 m_1 m_2}{n^2(n-1)}$$

Fisher's exact test (2)

Approximate chi-square test can be constructed using the following statistic:

$$\chi^2 = \left(\frac{a - E_0(A)}{\sqrt{Var_0(A)}} \right)^2$$

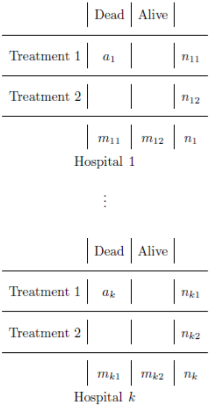
which asymptotically has chi-square distribution with 1 degree of freedom.

Remark

Obviously the square root of this statistic is asymptotically $N(0,1)$-distributed.

Mantel-Haenszel/log-rank test (1)

Let us generalize the situation and assume we have a sequence of 2x2 tables (say, k hospitals with two fixed treatments):



	Dead	Alive	
Treatment 1	a_1		n_{11}
Treatment 2			n_{12}
	m_{11}	m_{12}	n_1

Hospital 1

⋮

	Dead	Alive	
Treatment 1	a_k		n_{k1}
Treatment 2			n_{k2}
	m_{k1}	m_{k2}	n_k

Hospital k

Mantel-Haenszel/log-rank test (2)

For $i = 1, \ldots, k$, denote:

$p_{i1} = P\{\text{Dead} \mid \text{Treatment1, Hospital } i\}$ and $p_{i2} = P\{\text{Dead} \mid \text{Treatment2, Hospital } i\}$

The corresponding null hypothesis is now

$$H_0 : p_{11} = p_{12}, \ldots, p_{k1} = p_{k2}$$

and the following Mantel-Haenszel statistic can be used

$$MH = \frac{\sum_{i=1}^{k}(a_i - E_0(A_i))}{\sqrt{\sum_{i=1}^{k} Var_0(A_i)}}$$

Under H_0, the statistic MH is asymptotically $N(0,1)$ distributed

Mantel-Haenszel/log-rank test (3)

Question

How to apply this approach to censored survival data?

Idea

- combine the two samples (time points of failures/deaths for two populations)
- order the combined sample, obtain an ordered sample
- for each uncensored time point construct a 2x2 table similarly to previous
- compute MH statistic for this sequence of tables to test if $F_1 = F_2$

Mantel-Haenszel/log-rank test. Example (1)

Consider the following data:

Treatment Old	3, 5, 7, 9+, 18
Treatment New	12, 19, 20, 20+, 33+

Denote

- z – the combined ordered value (time point)
- n – total number of patients at risk (at time z)
- m_1 – number of patients who died (at time z)
- n_1 – number at risk in group Old (at time z)
- a – indicator if death is in group Old ($a = 1$) or New ($a = 0$)

trt	z	n	m_1	n_1	a	$E_0(A)$	r	$\frac{m_1(n-m_1)}{n-1}$	$\frac{n_1}{n}\left(1-\frac{n_1}{n}\right)$
Old	3	10	1	5	1	.50	.50	1	.2500
Old	5	9	1	4	1	.44	.56	1	.2469
Old	7	8	1	3	1	.38	.62	1	.2344
Old	9+		0		0				
New	12	6	1	1	0	.17	$-.17$	1	.1389
Old	18	5	1	1	1	.20	.80	1	.1600
New	19	4	1	0	0	0	0	1	0
New	20	3	1	0	0	0	0	1	0
New	20+								
New	33+								
Total					4	1.69	2.31		1.0302

Mantel-Haenszel/log-rank test. Example (2)

Now, under H_0 we have:

$$E_0 A = \frac{n_1 m_1}{n}$$

$$Var_0(A) = \frac{n_1 n_2 m_1 m_2}{n^2(n-1)} = \frac{m_1(n-m_1)}{n-1} \cdot \frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)$$

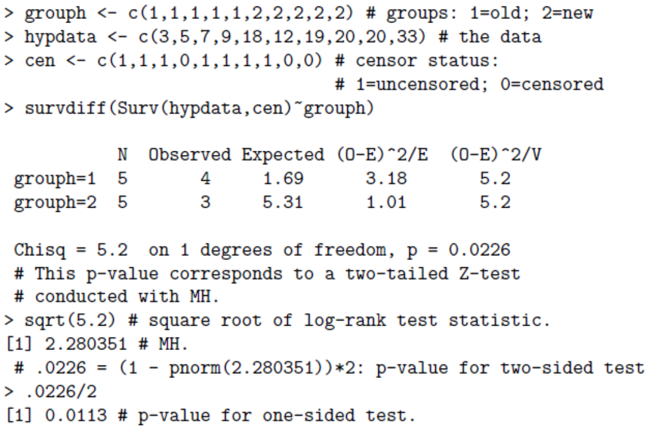
The corresponding *MH* statistic is

$$MH = \frac{\sum\left(a - \frac{n_1 m_1}{n}\right)}{\sqrt{\sum \frac{m_1(n-m_1)}{n-1} \cdot \frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)}} = \frac{2.31}{1.02} = 2.26$$

and using (one-tailed) Z-test, the p-value is 0.012.

Mantel-Haenszel/log-rank test. Example (3)

Solution to the same exercise in R is similar (although `survdiff` function gives chi-square statistic, the corresponding Z-statistic can be easily derived):



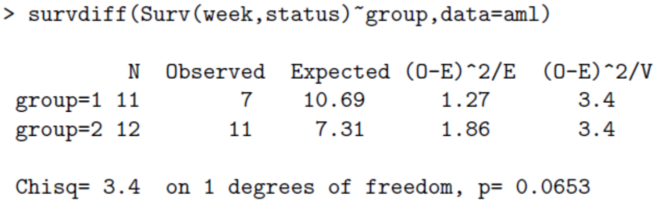
```
> grouph <- c(1,1,1,1,1,2,2,2,2,2) # groups: 1=old; 2=new
> hypdata <- c(3,5,7,9,18,12,19,20,20,33) # the data
> cen <- c(1,1,1,0,1,1,1,1,0,0) # censor status:
                                # 1=uncensored; 0=censored
> survdiff(Surv(hypdata,cen)~grouph)

          N  Observed Expected (O-E)^2/E  (O-E)^2/V
 grouph=1  5       4      1.69      3.18       5.2
 grouph=2  5       3      5.31      1.01       5.2

 Chisq = 5.2  on 1 degrees of freedom, p = 0.0226
 # This p-value corresponds to a two-tailed Z-test
 # conducted with MH.
> sqrt(5.2) # square root of log-rank test statistic.
[1] 2.280351 # MH.
 # .0226 = (1 - pnorm(2.280351))*2: p-value for two-sided test
> .0226/2
[1] 0.0113 # p-value for one-sided test.
```

Mantel-Haenszel test for AML example

The log-rank test on the AML data usin R:



```
> survdiff(Surv(week,status)~group,data=aml)

         N Observed Expected (O-E)^2/E (O-E)^2/V
group=1 11        7    10.69      1.27       3.4
group=2 12       11     7.31      1.86       3.4

 Chisq= 3.4  on 1 degrees of freedom, p= 0.0653
```

Thus, there is mild evidence to suggest that maintenance chemotherapy prolongs the remission period since the one-sided test is appropriate and its *p*-value is $0.0653/2 = .033$

Home assignment

Home assignment 4.
Observed survival times for two groups are following:

group 1:	1	1	3	4
group 2:	2	3+	5	5

The symbol (+) indicates censoring. Compare the survival curves in both groups with the help of log-rank test. Do the calculations both by yourself (using paper and pen and compare the result to chi-square critical value) and by using R (include both your commands and the results).