MAS3301 / MAS8311 Biostatistics
Part II: Survival

M. Farrow
School of Mathematics and Statistics
Newcastle University

Semester 2, 2009-10

5   Uncertainty in the Kaplan Meier Estimate

5.1   Introduction

In this section we will look at how to construct 'error bars' or confidence intervals for the Kaplan-Meier estimate of the survivor function. We will then go on to look at estimates and confidence intervals for the median survival time (and other quantiles). The following lemma plays an important role in the derivation of the confidence intervals.

A basic lemma

Lemma. *For a function g and random variable X we have*

$$
\begin{aligned}
E[g(X)] &\approx g[E(X)] \quad and, \\
\mathrm{var}[g(X)] &\approx \mathrm{var}(X)\{g'[E(X)]\}^2.
\end{aligned}
$$

In particular

$$
\mathrm{var}[\log(X)] \approx \frac{1}{\{E(X)\}^2}\mathrm{var}(X).
$$

Proof: The first two terms of the Taylor series for g are:

$$
g(x) = g(a) + (x - a)g'(a) + \cdots
$$

Setting $a = E(X)$ gives:

$$
g(X) \approx g[E(X)] + [X - E(X)]g'[E(X)]. \qquad (*)
$$

Taking expectations on both sides of $(*)$ gives

$$
E[g(X)] \approx g[E(X)].
$$

Squaring both sides of $(*)$ and taking the expectation gives

$$
\begin{aligned}
[g(X)]^2 &\approx g[E(X)]^2 + [X - E(X)]^2\{g'[E(X)]\}^2 \\
&\quad + 2g\,(E(X))\,[X - E(X)]g'[E(X)] \\
E\{[g(X)]^2\} &\approx \{g[E(X)]\}^2 + \mathrm{var}(X)\{g'[E(X)]\}^2
\end{aligned}
$$

so

$$
\begin{aligned}
\mathrm{var}[g(X)] &= E[g(X)^2] - E[g(X)]^2 \\
&\approx \mathrm{var}(X)\{g'[E(X)]\}^2.
\end{aligned}
$$

$\square$

5.2   Confidence intervals for the survivor function

In the KM estimate of the survivor function, the probability p_j of surviving the interval $[t_j, t_{j+1})$ is estimated by

$$
\hat{p}_j = 1 - d_j/N_j.
$$

If we know N_j, then the observed number d_j of deaths on the interval $[t_j, t_{j+1})$ has a binomial distribution $\mathrm{Bin}(N_j, p_j)$. We therefore have

$$
\mathrm{var}(\hat{p}_j \mid N_j) = \frac{N_j p_j(1 - p_j)}{N_j^2} \approx \frac{\hat{p}_j(1 - \hat{p}_j)}{N_j}
$$

Consider the log of the survivor function:

$$
\log[\hat{S}(t)] = \sum_{i=1}^{j} \log(\hat{p}_i), \quad \text{when } t_j \leq t < t_{j+1}. \tag{1}
$$

We now use the log variance formula in the lemma on both sides on the equation. First the left-hand side of (1) gives

$$\text{var}\{\log[\hat{S}(t)]\} \approx \frac{\text{var}[\hat{S}(t)]}{[\hat{S}(t)]^2}.$$

Applying the lemma to the right-hand of (1) gives

$$\text{var}[\log(\hat{p}_i)] \approx \frac{1}{[\text{E}(\hat{p}_i)]^2}\text{var}(\hat{p}_i) = \frac{1}{p_i^2}\frac{p_i(1-p_i)}{N_i} = \frac{1-p_i}{p_i N_i} \approx \frac{1-\hat{p}_i}{\hat{p}_i N_i}$$

so

$$\text{var}\left[\sum_{i=1}^{j}\log(\hat{p}_i)\right] \approx \sum_{i=1}^{j}\frac{1-\hat{p}_i}{\hat{p}_i N_i}.$$

Since var(LHS) of (1) equals var(RHS), we have

$$\frac{\text{var}[\hat{S}(t)]}{[\hat{S}(t)]^2} \approx \sum_{i=1}^{j}\frac{1-\hat{p}_i}{\hat{p}_i N_i} = \sum_{i=1}^{j}\frac{1-1+d_i/N_i}{(1-d_i/N_i)N_i} = \sum_{i=1}^{j}\frac{d_i}{N_i(N_i-d_i)}.$$

The standard error is

$$\text{s.e.}[\hat{S}(t)] \approx \hat{S}(t)\sqrt{\sum_{i=1}^{j}\frac{d_i}{N_i(N_i-d_i)}}$$

for $t_j \le t < t_{j+1}$.
This is known as *Greenwood's formula.*

When no censoring occurs, Greenwood's formula can be simplified. In fact, the variance can be shown to be the same as that calculated in Section 3.1, and Greenwood's formula becomes:

$$\text{s.e.}[\hat{S}(t)] = \hat{S}(t)\sqrt{\frac{1-\hat{S}(t)}{N_0\hat{S}(t)}}.$$

Note that this only applies if there is **no censoring** up to time t.

In order to construct an error bar for $\hat{S}(t)$ we need to make some sort of distributional assumption. The simplest is to assume that $\hat{S}(t)$ is normally distributed. Let $z_{\alpha/2}$ be such that $\Pr(Z > z_{\alpha/2}) = \alpha/2$, where $Z \sim N(0,1)$, then an approximate $100(1-\alpha)\%$ interval is given by

$$\hat{S}(t) \pm z_{\alpha/2} s[\hat{S}(t)]$$

where $s[\hat{S}(t)]$ is the standard error.

A snag with this, of course is that the distribution of $\hat{S}(t)$ is not really normal. Problems arise when $\hat{S}(t)$ is close to 0 or 1. One possible solution is to transform $\hat{S}(t)$ onto a $(-\infty, \infty)$ scale. For example, we could use a *logistic* transformation

$$W = \log\left(\frac{S}{1-S}\right)$$

or a *complementary log-log* transformation

$$W = \log[-\log(S)].$$

Consider the complementary log-log transformation. From above, we have

$$\mathrm{var}\{-\log[\hat{S}(t)]\} \approx \sum_{i=1}^{j} \frac{d_i}{N_i(N_i - d_i)}.$$

Applying the variance lemma with $X = -\log[\hat{S}(t)]$ we have

$$\mathrm{var}\left(\log\{-\log[\hat{S}(t)]\}\right) \approx \frac{1}{\{\log[\hat{S}(t)]\}^2} \sum_{i=1}^{j} \frac{d_i}{N_i(N_i - d_i)}$$

The standard error is the square root of this, say $\eta[\hat{S}(t)]$. Assuming W is normally distributed, a $100(1-\alpha)\%$ confidence interval for $\log\{-\log[S(t)]\}$ is

$$\log\{-\log[\hat{S}(t)]\} \pm z_{\alpha/2} \eta[\hat{S}(t)].$$

Now just transform back. A $100(1-\alpha)\%$ confidence interval for $\log[S(t)]$ is

$$\left(\log[\hat{S}(t)] \exp\{z_{\alpha/2}\eta[\hat{S}(t)]\}, \log[\hat{S}(t)] \exp\{-z_{\alpha/2}\eta[\hat{S}(t)]\}\right)$$

and a $100(1-\alpha)\%$ confidence interval for $S(t)$ is

$$\left([\hat{S}(t)]^{\exp\{z_{\alpha/2}\eta[\hat{S}(t)]\}}, [\hat{S}(t)]^{\exp\{-z_{\alpha/2}\eta[\hat{S}(t)]\}}\right).$$

Example: The data in the table are failure times, in days, of closely matched skin grafts on burn patients, from Woolson and Lachenbruch (1980) (slightly altered). Censored times are in brackets. The initial number $N_0 = 11$.

| | Failure time | | | | | $\hat{S}(t)$ |
j	t_j	N_j	c_j	d_j	$\hat{p}_j$	$(t_j \le t < t_{j+1})$
1	16	11	0	1	$(10/11) = 0.909$	0.909
2	18	10	0	1	$(9/10) = 0.900$	0.818
3	19	9	0	2	$(7/9) = 0.778$	0.636
4	22	7	0	1	$(6/7) = 0.857$	0.545
5	29	6	0	1	$(5/6) = 0.833$	0.455
6	37	5	2	1	$(4/5) = 0.800$	0.364
	(57)					
	(60)					
7	63	2	0	1	$(1/2) = 0.500$	0.182
8	93	1	0	1	$(0/1) = 0.000$	0.000

To find a 95% confidence interval for $S(25)$ we proceed as follows.
$\hat{S}(25) = 0.545$
Since there is no censoring before $t = 25$, we can use the simplified form of Greenwood's formula:

$$\text{s.e.}[\hat{S}(25)] = \hat{S}(25)\sqrt{\frac{1 - \hat{S}(25)}{N_0 \hat{S}(25)}} = 0.150$$

The confidence interval (assuming normality for $S(t)$) is $0.545 \pm 1.96 \times 0.150$ ie. $0.251 < S(25) < 0.840$.
Using the complementary log-log transformation,

$$\text{var}\left(\log\{-\log[\hat{S}(25)]\}\right) \approx \frac{1}{\{\log[\hat{S}(t)]\}^2} \sum_{i=1}^{5} \frac{d_i}{N_i(N_i - d_i)}$$

$$= \frac{0.0758}{\log(0.545)^2} = 0.2057$$

so the standard error is approx 0.4535.
A 95% CI for $\log\{-\log[S(25)]\}$ is $-0.4993 \pm 1.96 \times 0.4535$ ie. $-1.3882 < \log\{-\log[S(25)]\} < 0.3896$.
The corresponding CI for $S(25)$ is therefore $0.228 < S(25) < 0.779$.

5.3   Estimating the median and other percentiles

The median is arguably more useful than the mean with survival data because of the skewness. The usual nonparametric estimate of the median, when the estimated survivor function is a step function, is the smallest observed survival time for which the value of the estimated survivor function is less than or equal to 0.5. So, in the skin graft example, the estimate of the median survival time is 29 days.

Similarly we can estimate other percentiles. To find the 100α percentile (eg. $\alpha = 0.75$ for the upper quartile), you find the smallest observed survival time for which $\hat{S}$ is less than or equal to $(1-\alpha)$. Eg. For the upper quartile this is the minimum observed death time t_j such that $\hat{S}(t_j) \leq 0.25$.

We'd also like confidence intervals for percentiles. First recall that

$$\frac{d}{dt}S(t) = -f(t)$$

where f is the probability density function. Also recall the variance formula:

$$\text{var}[g(X)] \approx \text{var}(X)\{g'[E(X)]\}^2.$$

Putting $X = \hat{t}_\alpha$ and $g(x) = \hat{S}(x)$ gives

$$\text{var}[\hat{S}(\hat{t}_\alpha)] \approx \left[\frac{d\hat{S}(\hat{t}_\alpha)}{d\hat{t}_\alpha}\right]^2 \text{var}(\hat{t}_\alpha)$$

$$\approx [\hat{f}(\hat{t}_\alpha)]^2 \text{var}(\hat{t}_\alpha)$$

where $\hat{f}(\hat{t}_\alpha)$ is an estimate of the probability density function f of survival times at time t_α. Rearranging the equation gives

$$\text{var}(\hat{t}_\alpha) \approx [\hat{f}(\hat{t}_\alpha)]^{-2}\text{var}[\hat{S}(\hat{t}_\alpha)]$$

so the standard error of $\hat{t}_\alpha$ is approximately

$$\text{s.e.}[\hat{t}_\alpha] \approx \frac{1}{\hat{f}(\hat{t}_\alpha)}\text{s.e.}[\hat{S}(\hat{t}_\alpha)]$$

where $\text{s.e.}[\hat{S}(\hat{t}_\alpha)]$ is the standard error of $\hat{S}(\hat{t}_\alpha)$ which we estimate with Greenwood's formula. To finish constructing the CI we also need to estimate $f(\hat{t}_\alpha)$.

The probability density f is $-dS/dt$ so essentially we need an estimate of the gradient of S at $\hat{t}_\alpha$. Here is one way to do it: if you define t_- and t_+ to be two times with $t_- < t < t_+$ then an estimate of the gradient is

$$\frac{\hat{S}(t_+) - \hat{S}(t_-)}{t_+ - t_-}.$$

How do we pick t_- and t_+? A good way (even though it doesn't look nice) is as follows. Take

$$t_+ = \min\{t_j : \hat{S}(t_j) \leq (1-\alpha) - \epsilon\}$$
$$t_- = \max\{t_j : \hat{S}(t_j) \geq (1-\alpha) + \epsilon\}$$

for some fixed ϵ, say $\epsilon = 0.05$. The estimate of the standard error of $\hat{t}_\alpha$ is then given by

$$\text{s.e.}[\hat{t}_\alpha] \approx -\frac{t_+ - t_-}{\hat{S}(t_+) - \hat{S}(t_-)}\text{s.e.}[\hat{S}(\hat{t}_\alpha)].$$

Example Using the skin graft data from the last section, construct a confidence interval for the median.
$\hat{S}(29) = 0.455 \leq 0.5$ so $\hat{t}=29$.
There is no censoring up to $t = 29$, so we can use the simplified version of Greenwood's formula:

$$\text{s.e.}[\hat{S}(\hat{t})] \approx \hat{S}(\hat{t})\sqrt{\frac{1 - \hat{S}(\hat{t})}{N_0\hat{S}(\hat{t})}}$$

$$\approx 0.455\sqrt{\frac{1 - 0.455}{11 \times 0.455}} = 0.1501$$

Choose $\epsilon = 0.05$.

$$
\begin{aligned}
t_- &= \max\{t_j \text{ s.t. } \hat{S}(t_j) \geq 0.55\} = 19 \\
t_+ &= \min\{t_j \text{ s.t. } \hat{S}(t_j) \leq 0.45\} = 37
\end{aligned}
$$

So

$$
\hat{f}[\hat{t}] = \frac{\hat{S}(19) - \hat{S}(37)}{37 - 19} = \frac{0.636 - 0.364}{18} = 0.0151
$$

and

$$
\begin{aligned}
\text{s.e.}[\hat{t}] &\approx \frac{1}{\hat{f}(\hat{t})} \text{s.e.}[\hat{S}(\hat{t})] \\
&= \frac{0.1501}{0.0151} = 9.93
\end{aligned}
$$

Our approximate 95% confidence interval for the median is $29 \pm 1.96 \times 9.93$.
That is

$$
9.5 < t_{0.5} < 48.
$$

6 The Log-Rank Test

6.1 Introduction

Given two groups of people, we can ask the question: does one group tend to live longer than the other? A naïve approach is to plot the KM estimates for the two groups and check whether one survivor function is consistently higher than the other. If this is found to be the case, then it looks like one group does indeed live longer. However, this observation could be the result of chance variation – is there sufficient evidence in the two data sets to draw this conclusion? To be sure, we need a more formal hypothesis test.

Here's a brief informal recap on hypothesis testing. Recall that a hypothesis test has the following features. We have a (conservative) null hypothesis H_0 (which in this case is that the lifetime distributions are the same for the two groups), and an alternative hypothesis H_1 (that the two groups have different lifetime distributions). We then construct a statistic w that tends to be larger when the null hypothesis is false. Given the observed data we calculate w and compare it to

its distribution under the null: if it is in the tail of the distribution (ie. if the value of w is large) then we reject the null in favour of the alternative hypothesis.

6.2   The log-rank test

Suppose that we have two groups of survival data, Group 1 and Group 2. We start off by combining the death times of the two groups and construct intervals (just like we did for the KM estimator, but here we pool the two groups of death times). So we have death times $t_1, \ldots, t_m$ and intervals $[0, t_1), [t_1, t_2), \ldots$ etc. Let d_{1j} and d_{2j} be the number of deaths in the two groups on interval $[t_j, t_{j+1})$. Let n_{1j}, n_{2j} be the number of individuals in each group, known to be alive at the start of interval j (just as previously), so that $n_{1,j+1} = n_{1j} - c_{1,j} - d_{1,j}$ and similarly for Group 2. Define d_j and n_j by the sums:

$$d_j = d_{1j} + d_{2j}, \qquad n_j = n_{1j} + n_{2j}.$$

If the death times are not tied then one of d_{1j} and d_{2j} is 1 and the other is zero. For each interval j we can construct a 2×2 contingency table:

	Died	Did not die	
Group 1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
Group 2	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
	d_j	$n_j - d_j$	n_j

Consider the null hypothesis that there is no difference in the survival distribution of the two groups. One way to assessing this is to look at the extent of the difference between the observed numbers in each group that die on each interval and the number expected under the null. To do this we regard the marginal sums in the contingency table as fixed. Then the value of all the cells can be determined given only d_{1j}, so we consider this as our random variable. Under the null, the probability of a member of Group 1 dying at time t_j is the same as that for a member of Group 2, say p. Then the probability of observing d_{1j} deaths in group 1 and d_{2j} deaths in Group 2, given that there are d_j deaths at time t_j is:

$$\Pr(d_{1j} \text{ deaths in group 1} \mid d_j, n_{1j}, n_{2j}) =$$
$$\frac{\binom{n_{1j}}{d_{1j}} p^{d_{1j}}(1-p)^{n_{1j}-d_{1j}} \binom{n_{2j}}{d_{2j}} p^{d_{2j}}(1-p)^{n_{2j}-d_{2j}}}{\binom{n_j}{d_j} p^{d_j}(1-p)^{n_j-d_j}} = \frac{\binom{n_{1j}}{d_{1j}} \binom{n_{2j}}{d_{2j}}}{\binom{n_j}{d_j}}$$

Notice that p cancels out.

It can be shown that the mean of d_{1j} is

$$e_{1j} = \frac{n_{1j}d_j}{n_j}.$$

If you're interested in how to do this, d_{1j} has a hypergeometric distribution and you use generating functions to calculate its moments – see Grimmet and Stirzaker's book, Probability and Random Processes. This expectation makes sense: it corresponds to sharing the deaths out in proportion to how many people there are in each group. Similarly you can show the variance is:

$$v_{1j} = \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}.$$

These formulae can be derived in a more straight forward way in the case $d_j = 1$ ie. without recourse to generating functions.

Next we combine the information from each of the time intervals to give an overall measure of how much the observed d_{1j} deviate from their expectations.

Define U by
$$U = \sum_1^m (d_{1j} - e_{1j}) = \sum_1^m d_{1j} - \sum_1^m e_{1j} = O_1 - E_1$$

Under the null this will have mean zero, since the expectation of d_{1j} is e_{1j}. If we assume U is normally distributed (sum of random variables, central limit theorem), then $U \sim N(0, V)$ where

$$V = \sum_1^m v_{1j}.$$

So $[U/\sqrt{(V)}] \sim N(0, 1)$ and
$$W = \frac{U^2}{V} \sim \chi_1^2.$$

To test the null hypothesis, we compute the statistics W and compare it to the upper tail of a χ_1^2 distribution. This is the *log-rank* test.

6.3   Example

Data gathered from a trial of a new drug, looking at time until remission (in weeks) in patients with leukemia.

Control: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12

Drug: 6, 6, 6, 6*, 7, 9*, 10*, 11, 11*, 16, 17*, 19*, 20*, 25, 32*, 32*

The leukemia data give the following table:

t_j	d_{1j}	d_{2j}	d_j	n_{1j}	n_{2j}	n_j	e_{1j}	v_{1j}
1	2	0	2	17	16	33	34/33	0.483
2	2	0	2	15	16	31	30/31	0.483
3	1	0	1	13	16	29	13/29	0.247
4	2	0	2	12	16	28	24/28	0.472
5	2	0	2	10	16	26	20/26	0.454
6	0	3	3	8	16	24	24/24	0.609
7	0	1	1	8	12	20	8/20	0.240
8	4	0	4	8	11	19	32/19	0.813
11	2	1	3	4	9	13	12/13	0.531
12	2	0	2	2	7	9	4/9	0.302
16	0	1	1	0	7	7	0/7	0.0
25	0	1	1	0	3	3	0/3	0.0
	O_1						E_1	V

This gives $O_1 = 17$, $E_1 = 8.5$, and $V = 4.634$. It follows that the test statistic is

$$W = \frac{(17 - 8.5)^2}{4.634} = 15.6$$

This is highly significant against χ_1^2 so we reject the null hypothesis that the distribution of survival times are the same in the two groups.

6.4 Simplified form

An alternative simpler statistic is:

$$W^* = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

where $O_2 = \sum_1^m d_{2j}$ and $E_2 = \sum_1^m e_{2j}$ with $e_{2j} = n_{2j}d_j/n_j$.

This is also compared with χ_1^2 and is usually an adequate approximation though it errs on the conservative side. The advantage is that it's easier to compute.

Note that

$$O_1 + O_2 = \sum(d_{1j} + d_{2j}) = \sum d_j, \quad \text{and}$$

$$E_1 + E_2 = \sum \left(\frac{n_{1j}d_j}{n_j} + \frac{n_{2j}d_j}{n_j} \right) = \sum d_j$$

ie. $O_1 + O_2 = E_1 + E_2$.

This makes some calculations a bit quicker.

Example: Carry out the simplified test for the following data

Group 1: 10, 11*, 12*, 14, 16*, 22
Group 2: 13, 22, 26*, 29, 35*, 40*

t_j	d_{1j}	d_{2j}	d_j	n_{1j}	n_{2j}	n_j	e_{1j}
10	1	0	1	6	6	12	6/12
13	0	1	1	3	6	9	3/9
14	1	0	1	3	5	8	3/8
22	1	1	2	1	5	6	2/6
29	0	1	1	0	3	3	0
	$O_1 = 3$	$O_2 = 3$					$E_1 = 1.54$

$E_2 = O_1 + O_2 - E_1 = 3 + 3 - 1.54 = 4.46$ so

$$W^* = \frac{(3 - 1.54)^2}{1.54} + \frac{(3 - 4.46)^2}{4.46} = 1.86$$

This is not significant against χ_1^2.

6.5 Relative rates

If we reject the null hypothesis, we want to know in which group the survival times are shorter. In order to do this we compute the *relative death rates* $R_1 = O_1/E_1$ and $R_2 = O_2/E_2$. One group will have $R > 1$ and the other $R < 1$: survival times are shorter in the group with $R > 1$. If we assume that the underlying hazard functions for the two groups are proportional ie.

$$h_1(t) = \psi h_2(t)$$

then an estimate of the hazard ratio is

$$\psi = \frac{O_1/E_1}{O_2/E_2}.$$

7 More on Log-Rank Tests

In this section we look at how the log-rank test can be extended to handle more interesting situations.

7.1 The log-rank test for multiple groups

Suppose we have r groups of individuals, with $r > 2$. We're interested in testing whether all the groups are the same ie. the null hypothesis is that all the groups have the same survival distribution. Our approach and notation is the same as previously: we pool all the death times together to define intervals $[0, t_1), [t_1, t_2)$, etc. We then have d_{kj} deaths in group k on interval j and n_{kj} individuals alive and uncensored from group k at the start of interval j.

We can extend the simplified form of the log-rank test quite readily.

Define a new test statistic
$$W^* = \sum_{k=1}^{r} \frac{(O_k - E_k)^2}{E_k}$$
where

$$O_k = \sum_{j=1}^{m} d_{kj} \quad \text{is the observed number of deaths in group } k$$

$$E_k = \sum_{j=1}^{m} e_{kj} \quad \text{is the expected number of deaths in group } k$$

with

$$e_{kj} = \frac{n_{kj} d_j}{n_j}.$$

This time there are $r-1$ degrees of freedom. (To convince yourself of this, think of a contingency table on each interval like the one described in the last section: it has $2 \times r$ cells and $r+1$ constraints.) We therefore compare W^* to the critical points of the χ^2_{r-1} distribution.

The more formal version of the test can also be extended to deal with more than 2 groups, but for the purposes of this course we will only use the simplified form.

7.2   The stratified log-rank test

Sometimes we wish to compare two or more groups but taking into account another variable eg. age group, sex, centre. Consider the following example. Suppose we're conducting a clinical trial of two different treatments for a disease, but we know that men and women will probably have different responses to the treatments. In this example there are two groups (2 different treatments) and two strata (male / female). Because we think men and women respond differently we don't want to compare all of group 1 with all of group 2. Instead we want to compare the two groups of men and the two groups of women separately. Essentially we calculate the U and V statistics for each stratum, and combine these to give an overall statistic. The null hypothesis is that there is no difference between the groups in the survival distribution for each stratum.

Suppose we have $k = 1, \ldots, r$ groups (think of these are 'treatments') and $l = 1, \ldots, s$ strata (think of these as age ranges or some other division of each group). Working within each stratum separately, pool the death times together as previously to define time intervals $j = 1, \ldots, m_l$. For the l-th stratum this defines

$$d_{lkj} = \text{number of deaths in stratum } l, \text{ group } k, \text{ interval } j$$

and n_{lkj}, e_{lkj} similarly. Modifying the simpler form of the log-rank test, we define

$$W^* = \sum_{k=1}^{r} \frac{\left[\sum_{l=1}^{s}(O_{lk} - E_{lk})\right]^2}{\sum_{l=1}^{s} E_{lk}}$$

where

$$O_{lk} = \sum_{j=1}^{m_l} d_{lkj} = \text{observed number of deaths in stratum } l, \text{ group } k,$$

$$E_{lk} = \sum_{j=1}^{m_l} e_{lkj} = \text{expected number of deaths in stratum } l, \text{ group } k,$$

and

$$e_{lkj} = \frac{n_{lkj} d_{lj}}{n_{lj}}.$$

Like the grouped test, we compare W^* to the critical points of a χ^2_{r-1} distribution. In effect, the test can be carried out via the following steps:

- Split the data up according to the strata.

- Prepare one table for each stratum $l = 1, \ldots, s$.

- Calculate O_{lk} and E_{lk} for each table.

- Combine the results from each table into the test statistic.

7.3   Example – Stratified Test

This example is rather small but it illustrates the simple version of the stratified test. The data, given by Lee (1992), refer to a study by the University of Oklahoma Health Sciences Centre. They are survival times for patients with melanoma (a skin cancer). In each case the tumour was surgically removed then patients in Group 1 were given BCG vaccine while those in Group 2 were given *c.parvum* vaccine. The patients were stratified by age-group: 21-40, 41-60 and 61-.

	21-40		41-60		61-	
Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
19	27*	34*	8	10	25*
24*	21*	4	11*	5	8
8	18*	17*	23*		11*
17*	16*		12*		
17*	7		15*		
34*	12*		8*		
	24		8*		
	8				
	8*				

Stratum 1: 21-40

t_j	n_{1j}	n_{2j}	d_{1j}	d_{2j}	n_j	d_j	e_{1j}	e_{2j}
7	6	9	0	1	15	1	0.400	0.600
8	6	8	1	1	14	2	0.857	1.143
19	3	3	1	0	6	1	0.500	0.500
24	2	2	0	1	4	1	0.500	0.500
			2	3			2.257	2.743
			O_{11}	O_{21}			E_{11}	E_{21}

Stratum 2: 41-60

t_j	n_{1j}	n_{2j}	d_{1j}	d_{2j}	n_j	d_j	e_{1j}	e_{2j}
4	3	7	1	0	10	1	0.300	0.700
8	2	7	0	1	9	1	0.222	0.778
			1	1			0.522	1.478
			O_{12}	O_{22}			E_{12}	E_{22}

Stratum 3: 61-

t_j	n_{1j}	n_{2j}	d_{1j}	d_{2j}	n_j	d_j	e_{1j}	e_{2j}
5	2	3	1	0	5	1	0.400	0.600
8	1	3	0	1	4	1	0.250	0.750
10	1	2	1	0	3	1	0.333	0.667
			2	1			0.983	2.017
			O_{13}	O_{23}			E_{13}	E_{23}

$$\sum_{s=1}^{3} O_{1s} = 2 + 1 + 2 = 5$$

$$\sum_{s=1}^{3} O_{2s} = 3 + 1 + 1 = 5$$

$$\sum_{s=1}^{3} E_{1s} = 2.257 + 0.522 + 0.983 = 3.762$$

$$\sum_{s=1}^{3} E_{2s} = 2.743 + 1.478 + 2.017 = 6.238$$

$$W^{\star} = \frac{(5 - 3.762)^2}{3.762} + \frac{(5 - 6.238)^2}{6.238} = 0.4074 + 0.2457 = 0.653.$$

Compare 0.653 with χ_1^2. The result is not significant so we do not reject the null hypothesis that choice of treatment has no effect on survival.

7.4 The log-rank test for trend

Sometimes when we have survival data for different groups, these groups are ordered in some way eg. by age group, increasing doses of treatment, or stage of a disease at which the patient was

first diagnosed. It might be that there is no significant difference between the groups as a whole (using the standard grouped log-rank test above), but when we introduce the ordering there is a significant trend. In order to test for trend, we introduce a weighting w_k, $k = 1, \ldots, r$, for each of the groups. Typically you might take $w_k = k$, but uneven weightings can be used – usually you'd like some justification from the particular application being studied.

To make such a test, we first split time up into intervals in the usual way by pooling the death times from all the groups. Next define the statistics

$$U_k = \sum_{j=1}^{m} w_k(d_{kj} - e_{kj}) \qquad (**)$$

$$= w_k(O_k - E_k)$$

for each group $k = 1, \ldots, r$. Then define

$$U_T = \sum_k U_k.$$

(The 'T' is for 'trend').

Each U_k measures how much group k differs from its expected number of deaths.

The variance of U_T is

$$V_T = \sum_{k=1}^{r} (w_k - \bar{w})^2 E_k$$

where

$$\bar{w} = \frac{\sum_k w_k E_k}{\sum_k E_k}.$$

The statistic U_T^2/V_T is approximately distributed as χ_1^2 under the hypothesis of no trend across the groups.

7.5 Example – Test for Trend

- Group 1 times: $8, 17^*, 17^*, 19, 24^*, 34^*$

- Group 2 times: $4, 17^*, 34^*$

- Group 3 times: $5, 10$

We'll use weights: $w_1 = -1$, $w_2 = 0$, and $w_3 = 1$.

t_j	d_{1j}	d_{2j}	d_{3j}	n_{1j}	n_{2j}	n_{3j}	e_{1j}	e_{2j}	e_{3j}
4	0	1	0	6	3	2	6/11	3/11	2/11
5	0	0	1	6	2	2	6/10	2/10	2/10
8	1	0	0	6	2	1	6/9	2/9	1/9
10	0	0	1	5	2	1	5/8	2/8	1/8
19	1	0	0	3	1	0	3/4	1/4	0
Total	2	1	2				3.18	1.19	0.62
	O_1	O_2	O_3				E_1	E_2	E_3

Then

$$U_T = \sum w_k(O_k - E_k) = -1 \times (2 - 3.18) + 1 \times (2 - 0.62) = 2.56$$

and

$$\bar{w} = \frac{\sum w_k E_k}{\sum E_k} = \frac{0.62 - 3.18}{5} = -0.51.$$

The variance term is

$$V_T = \sum (w_k - \bar{w})^2 E_k$$
$$= (-1 + 0.51)^2 \times 3.18 + 0.51^2 \times 1.19 + (1 + 0.51)^2 \times 0.62 = 2.49.$$

Finally,

$$W_T = U_T^2/V_T = 2.65.$$

Compared to χ_1^2 this has a p-value of 0.103 – so there is weak evidence of a trend.

Tutorial Examples 2

1. Find the survivor function associated with the following hazard function:

$$h(t) = \frac{t^2}{1 + \lambda^2 t^2}.$$

For what values of λ is this a valid survivor function?

2. Data are given below relating to a clinical trial of a new radiotherapy treatment for cervical cancer. The new treatment was tested against a control treatment. The grade of the cancer at the start of treatment was recorded for each individual: it was either 1 (early) or 2 (advanced). The clinicians conducting the trial were interested in determining whether the survival experiences under the two treatments were different, adjusting for the grade of the cancer at the start of treatment. Apply a suitable test.

Time	Status	Group	Stage
90	1	Control	2
150	1	Control	2
291	1	Control	1
362	1	Treatment	2
373	1	Treatment	2
383	0	Treatment	2
468	0	Control	1
519	0	Treatment	2
563	0	Treatment	2
650	0	Treatment	2
680	1	Control	2
827	1	Treatment	2
837	1	Control	2
890	0	Control	1
919	0	Treatment	1
978	0	Treatment	1
1090	0	Control	1
1100	0	Treatment	1
1113	0	Control	2
1153	1	Control	1
1297	1	Control	2
1307	1	Treatment	1
1360	0	Treatment	1
1429	1	Control	1
1476	0	Treatment	1
1577	0	Control	2

3. Figures 1-3 show plots of the estimated survivor function for a set of survival data. Each figure plots the same survivor function but with different axes, as labelled. Use the graphs to pick a possible parametric model for the data, and write down rough estimates of the model parameters.

4. Calculate the Kaplan Meier estimate of the survivor function for the following survival data. A "*" denotes a right-censored time.

$$3, 6, 10^*, 15, 33, 34^*, 38, 40, 40^*, 49, 49, 50^*, 54^*, 56^*$$

Assuming $\hat{S}$ is normally distributed, obtain a 99% confidence interval for $\hat{S}(12)$. (The 99.5th percentile for the standard normal distribution is at 2.58 standard deviations from the mean.) Comment on the interval you have obtained.
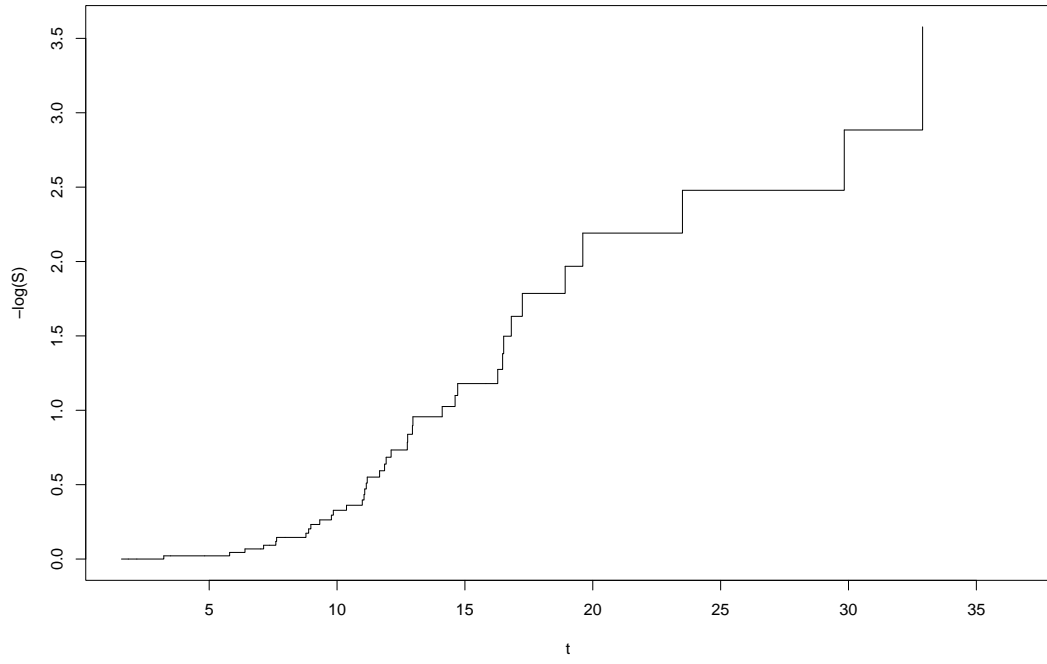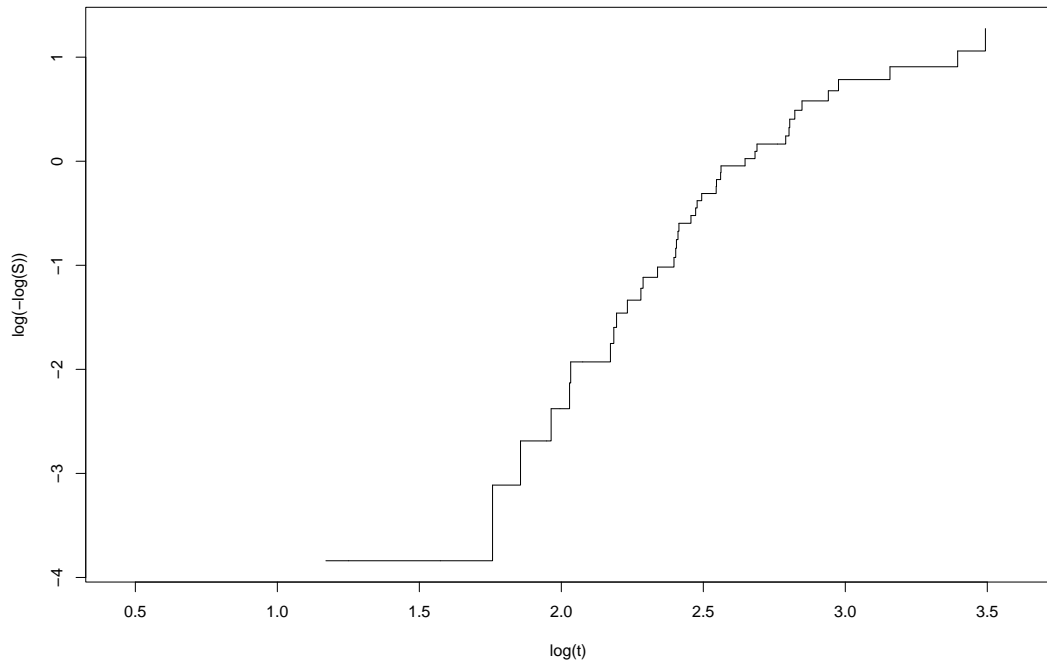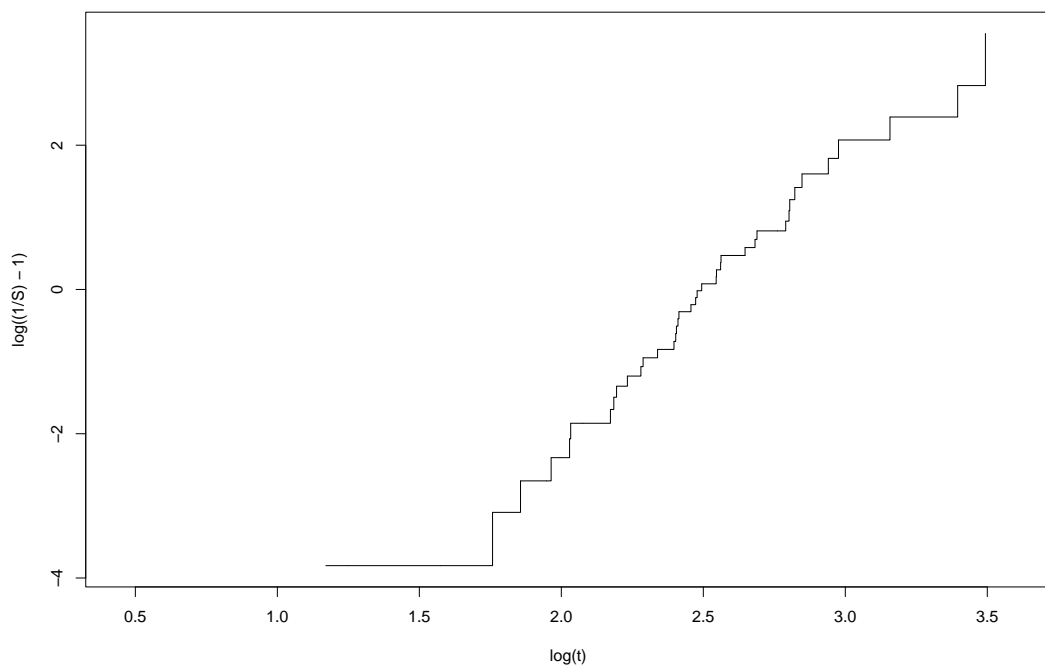
Figure 8: $-\log(S)$ vs t

Figure 9: $\log(-\log(S))$ vs $\log(t)$

Figure 10: $\log(\frac{1}{S} - 1)$ vs $\log(t)$