KM formula = product limit formula

$$\hat{S}(t_{(j-1)}) = \prod_{i=1}^{j-1} \hat{\Pr}(T > t_{(i)} \mid T \geq t_{(i)})$$

The above KM formula can also be expressed as a product limit if we substitute for the survival probability $\hat{S}(t_{(j-1)})$, the product of all fractions that estimate the conditional probabilities for failure times $t_{(j-1)}$ and earlier.

EXAMPLE

$$\hat{S}(10) = .8067 \times \frac{14}{15} = .7529$$
$$= \boxed{\frac{18}{21} \times \frac{16}{17}} \times \frac{14}{15}$$

$$\hat{S}(16) = .6902 \times \frac{10}{11}$$
$$= \boxed{\frac{18}{21} \times \frac{16}{17} \times \frac{14}{15} \times \frac{11}{12}} \times \frac{10}{11}$$

For example, the probability of surviving past ten weeks is given in the table for group 1 (page 55) by .8067 times 14/15, which equals .7529. But the .8067 can be alternatively written as the product of the fractions 18/21 and 16/17. Thus, the product limit formula for surviving past 10 weeks is given by the triple product shown here.

Similarly, the probability of surviving past sixteen weeks can be written either as .6902 × 10/11, or equivalently as the five-way product of fractions shown here.

$$\hat{S}(t_{(j)}) = \prod_{i=1}^{j} \hat{\Pr}[T > t_{(i)} \mid T \geq t_{(i)}]$$
$$= \hat{S}(t_{(j-1)}) \times \hat{\Pr}(T > t_{(j)} \mid T \geq t_{(j)})$$

The general expression for the product limit formula for the KM survival estimate is shown here together with the general KM formula given earlier. Both expressions are equivalent.

Math proof:

Pr(A and B) = Pr(A) × Pr(B | A) always

A simple mathematical proof of the KM formula can be described in probability terms. One of the basic rules of probability is that the probability of a joint event, say A and B, is equal to the probability of one event, say A, times the conditional probability of the other event, B, given A.

A = "$T \geq t_{(j)}$" → A and B = B
B = "$T > t_{(j)}$"

Pr(A and B) = Pr(B) = $S(t_{(j)})$

If we let A be the event that a subject survives to at least time $t_{(j)}$ and we let B be the event that a subject survives past time $t_{(j)}$, then the joint event A and B simplifies to the event B, which is inclusive of A. It follows that the probability of A and B equals the probability of surviving past time $t_{(j)}$.

No failures during $t_{(j-1)} < T < t_{(j)}$

Pr(A) = Pr($T > t_{(j-1)}$) = $S(t_{(j-1)})$

Also, because $t_{(j)}$ is the next failure time after $t_{(j-1)}$, there can be no failures after time $t_{(j-1)}$ and before time $t_{(j)}$. Therefore, the probability of A is equivalent to the probability of surviving past the (j – 1)th ordered failure time.

$\Pr(B \mid A) = \Pr(T > t_{(j)} \mid T \geq t_{(j)})$

Furthermore, the conditional probability of B given A is equivalent to the conditional probability in the KM formula.

Thus, from Pr(A and B) formula,

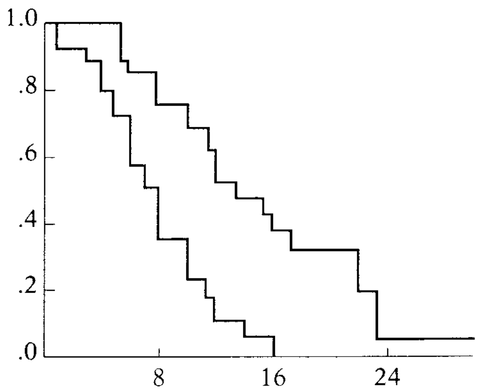
$S(t_{(j)}) = S(t_{(j-1)}) \times \Pr(T > t_{(j)} \mid T \geq t_{(j)})$

Thus, using the basic rules of probability, the KM formula can be derived.

IV. The Log–Rank Test for Two Groups

We now describe how to evaluate whether or not KM curves for two or more groups are statistically equivalent. In this section we consider two groups only. The most popular testing method is called the log–rank test.

Are KM curves statistically equivalent?



When we state that two KM curves are "statistically equivalent," we mean that, based on a testing procedure that compares the two curves in some "overall sense," we do not have evidence to indicate that the true (population) survival curves are different.

- Chi-square test
- Overall comparison of KM curves
- Observed versus expected counts
- Categories defined by ordered failure times

The log–rank test is a large-sample chi-square test that uses as its test criterion a statistic that provides an overall comparison of the KM curves being compared. This (log–rank) statistic, like many other statistics used in other kinds of chi-square tests, makes use of observed versus expected cell counts over categories of outcomes. The categories for the log–rank statistic are defined by each of the ordered failure times for the entire set of data being analyzed.

EXAMPLE

Remission data: $n = 42$

	# failures		# in risk set	
$t_{(j)}$	m_{1j}	m_{2j}	n_{1j}	n_{2j}
1	0	2	21	21
2	0	2	21	19
3	0	1	21	17
(4)	0	2	21	16
5	0	2	21	14
6	3	0	21	12
7	1	0	17	12
8	0	4	16	12
(10)	1	0	15	8
11	0	2	13	8
12	0	2	12	6
13	1	0	12	4
15	0	1	11	4
16	1	0	11	3
17	0	1	10	3
22	1	1	7	2
23	1	1	6	1

As an example of the information required for the log–rank test, we again consider the comparison of the treatment (group 1) and placebo (group 2) subjects in the remission data on 42 leukemia patients.

Here, for each ordered failure time, $t_{(j)}$, in the entire set of data, we show the numbers of subjects (m_{ij}) failing at that time, separately by group (i), followed by the numbers of subjects (n_{ij}) in the risk set at that time, also separately by group.

Thus, for example, at week 4, no subjects failed in group 1, whereas two subjects failed in group 2. Also, at week 4, the risk set for group 1 contains 21 persons, whereas the risk set for group 2 contains 16 persons.

Similarly, at week 10, one subject failed in group 1, and no subjects failed at group 2; the risk sets for each group contain 15 and 8 subjects, respectively.

Expected cell counts:

$$e_{1j} = \left(\frac{n_{1j}}{n_{1j} + n_{2j}}\right) \times \left(m_{1j} + m_{2j}\right)$$

↑ Proportion in risk set

↑ # of failures over both groups

$$e_{2j} = \left(\frac{n_{2j}}{n_{1j} + n_{2j}}\right) \times \left(m_{1j} + m_{2j}\right)$$

We now expand the previous table to include expected cell counts and observed minus expected values for each group at each ordered failure time. The formula for the expected cell counts is shown here for each group. For group 1, this formula computes the expected number at time j (i.e., e_{1j}) as the proportion of the total subjects in both groups who are at risk at time j, that is, $n_{1j}/(n_{1j} + n_{2j})$, multiplied by the total number of failures at that time over both groups (i.e., $m_{1j} + m_{2j}$). For group 2, e_{2j} is computed similarly.

EXAMPLE

Expanded Table (Remission Data)

		# failures		# in risk set		# expected		Observed–expected	
j	$t_{(j)}$	m_{1j}	m_{2j}	n_{1j}	n_{2j}	e_{1j}	e_{2j}	$m_{1j}-e_{1j}$	$m_{2j}-e_{2j}$
1	1	0	2	21	21	(21/42) × 2	(21/42) × 2	−1.00	1.00
2	2	0	2	21	19	(21/40) × 2	(19/40) × 2	−1.05	1.05
3	3	0	1	21	17	(21/38) × 1	(17/38) × 1	−0.55	0.55
4	4	0	2	21	16	(21/37) × 2	(16/37) × 2	−1.14	1.14
5	5	0	2	21	14	(21/35) × 2	(14/35) × 2	−1.20	1.20
6	6	3	0	21	12	(21/33) × 3	(12/33) × 3	1.09	−1.09
7	7	1	0	17	12	(17/29) × 1	(12/29) × 1	0.41	−0.41
8	8	0	4	16	12	(16/28) × 4	(12/28) × 4	−2.29	2.29
9	10	1	0	15	8	(15/23) × 1	(8/23) × 1	0.35	−0.35
10	11	0	2	13	8	(13/21) × 2	(8/21) × 2	−1.24	1.24
11	12	0	2	12	6	(12/18) × 2	(6/18) × 2	−1.33	1.33
12	13	1	0	12	4	(12/16) × 1	(4/16) × 1	0.25	−0.25
13	15	0	1	11	4	(11/15) × 1	(4/15) × 1	−0.73	0.73
14	16	1	0	11	3	(11/14) × 1	(3/14) × 1	0.21	−0.21
15	17	0	1	10	3	(10/13) × 1	(3/13) × 1	−0.77	0.77
16	22	1	1	7	2	(7/9) × 2	(2/9) × 2	−0.56	0.56
17	23	1	1	6	1	(6/7) × 2	(1/7) × 2	−0.71	0.71
Totals		9	21			19.26	10.74	−10.26	10.26

# of failure times

↓

$$O_i - E_j = \sum_{j=1}^{17}\left(m_{ij} - e_{ij}\right), \qquad i = 1, 2$$

EXAMPLE

$O_1 - E_1 = -10.26$

$O_2 - E_2 = 10.26$

When two groups are being compared, the log–rank test statistic is formed using the sum of the observed minus expected counts over all failure times for one of the two groups. In this example, this sum is −10.26 for group 1 and 10.26 for group 2. We will use the group 2 value to carry out the test, but as we can see, except for the minus sign, the difference is the same for the two groups.

Two groups:
$O_2 - E_2$ = summed observed minus expected score for group 2

$$\text{Log-rank statistic} = \frac{(O_2 - E_2)^2}{\text{Var}(O_2 - E_2)}$$

For the two-group case, the log–rank statistic, shown here at the left, is computed by dividing the square of the summed observed minus expected score for one of the groups—say, group 2—by the estimated variance of the summed observed minus expected score.

$$\text{Var}(O_i - E_i) = \sum_j \frac{n_{1j} n_{2j} (m_{1j} + m_{2j})(n_{1j} + n_{2j} - m_{1j} - m_{2j})}{(n_{1j} + n_{2j})^2 (n_{1j} + n_{2j} - 1)}$$

$i = 1, 2$

The expression for the estimated variance is shown here. For two groups, the variance formula is the same for each group. This variance formula involves the number in the risk set in each group (n_{ij}) and the number of failures in each group (m_{ij}) at time j. The summation is over all distinct failure times.

H_0: no difference between survival curves

Log–rank statistic $\sim \chi^2$ with 1 df under H_0

The null hypothesis being tested is that there is no overall difference between the two survival curves. Under this null hypothesis, the log–rank statistic is approximately chi-square with one degree of freedom. Thus, a P-value for the log–rank test is determined from tables of the chi-square distribution.

Computer programs:
SPIDA's km:
- descriptive statistics for KM curves
- log–rank statistic
- Peto statistic

SAS's lifetest

Several computer programs are available for calculating the log–rank statistic. For example, the **SPIDA** package has a procedure called "**km**" that computes descriptive information about Kaplan–Meier curves, the log–rank statistic, and an alternative statistic called the Peto statistic, to be described later. Other packages, like **SAS** and **BMDP**, have procedures that provide results similar to those of SPIDA. A comparison of SPIDA, SAS, and BMDP procedures and output is provided in Appendix A at the back of this text.

EXAMPLE

Using SPIDA: Remission Data

Group	Size	%Cen	LQ	Med	UQ	0.95 Med CI	
0	21	57.148	13	23		13.000	
1	21	0.000	4	8	12	4.000	8

df:1 Log–rank: 16.793, P-value: 0.0, Peto: 9.954, P-value: 0.002

For the remission data, the printout from using the SPIDA "km" procedure is shown here. The log–rank statistic is 16.793 and the corresponding P-value is zero to three decimal places. This P-value indicates that the null hypothesis should be rejected. We can therefore conclude that the treatment and placebo groups have significantly different KM survival curves.

EXAMPLE

$O_2 - E_2 = 10.26$

$\widehat{\text{Var}}(O_2 - E_2) = 6.2685$

$$\text{Log–rank statistic} = \frac{(O_2 - E_2)^2}{\widehat{\text{Var}}(O_2 - E_2)} = \frac{(10.26)^2}{6.2685} = 16.793$$

Although the use of a computer is the preferred way to calculate the log–rank statistic, we provide here some of the details of the calculation. We have already seen from earlier computations that the value of $O_2 - E_2$ is 10.26. The estimated variance of $O_2 - E_2$ is computed from the variance formula above to be 6.2685. The log–rank statistic then is obtained by squaring 10.26 and dividing by 6.285, which yields 16.793, as shown on the computer printout.

Approximate formula:

$$X^2 = \sum_{i}^{\#\text{ of groups}} \frac{(O_i - E_i)^2}{E_i}$$

An approximation to the log–rank statistic, shown here, can be calculated using observed and expected values for each group without having to compute the variance formula. The approximate formula is of the classic chi-square form that sums over each group being compared the square of the observed minus expected value divided by the expected value.

EXAMPLE

$$X^2 = \frac{(-10.26)^2}{19.26} + \frac{(10.26)^2}{10.74}$$

$= 15.276$ (conservative)

Log–rank statistic $= 16.793$

The calculation of the approximate formula is shown here for the remission data. The expected values are 19.26 and 10.74 for groups 1 and 2, respectively. The chi-square value obtained is 15.276, which is slightly smaller than the log–rank statistic of 16.793. Thus, for these data, the approximate formula provides a more conservative test—i.e., it is less likely to reject the null hypothesis—than the (exact) log–rank statistic for these data.

V. The Log–Rank Test for Several Groups

H_0: All survival curves are the same.

The log–rank test can also be used to compare three or more survival curves. The null hypothesis for this more general situation is that all survival curves are the same.

Log–rank statistics for > 2 groups involves variances and covariances of $O_i - E_i$.

Matrix formula: See Appendix at end of this chapter.

Although the same tabular layout can be used to carry out the calculations when there are more than two groups, the test statistic is more complicated mathematically, involving both variances and covariances of summed observed minus expected scores for each group. A convenient mathematical formula can be given in matrix terms. We present the matrix formula for the interested reader in an Appendix at the end of this chapter.