Use computer program for calculations.

We will not describe further details about the calculation of the log–rank statistic, because a computer program can easily carry out the computations from the basic data file. Instead, we illustrate the use of this test with data involving more than two groups.

G (≥ 2) groups:
log–rank statistic $\sim\chi^2$ with $G-1$ df

If the number of groups being compared is G (≥ 2), then the log–rank statistic has approximately a large sample distribution with $G - 1$ degrees of freedom. Therefore, the decision about significance is made using chi-square tables with the appropriate degrees of freedom.

Approximation formula:

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

Not required because computer program calculates the exact log–rank statistic

The approximate formula previously described involving only observed and expected values without variance or covariance calculations can also be used when there are more than two groups being compared. However, practically speaking, the use of this approximate formula is not required as long as a computer program is available to calculate the exact log–rank statistic.

We now provide an example to illustrate the use of the log–rank statistic to compare more than two groups.

EXAMPLE

vets.dat: survival time in days,

$n = 137$

Veteran's Administration Lung Cancer Trial

Column 1: Treatment standard = 1, test = 2)
Column 2: Cell type 1 (large = 1, other = 0)
Column 3: Cell type 2 (adeno = 1 other = 0)
Column 4: Cell type 3 (small = 1, other = 0)
Column 5: Cell type 4 (squamous = 1, other = 0)
Column 6: Survival time (days)
Column 7: Performance Status
(0 = worst, ..., 100 = best)
Column 8: Disease duration (months)
Column 9: Age
Column 10: Prior therapy (none = 0, some = 1)
Column 11: Status (0 = censored, 1 = died)

The data set "vets.dat" considers survival times in days for 137 patients from the Veteran's Administration Lung Cancer Trial cited by Kalbfleisch and Prentice in their text (*The Statistical Analysis of Survival Time Data*, John Wiley, pp. 223–224, 1980). Failure status is defined by the status variable (column 11). A complete list of the variables as stored in a SPIDA file is shown here; the actual data set is provided in an appendix.

Among the variables listed, we now focus on the performance status variable (column 7). This variable is an interval variable, so before we can obtain KM curves and the log–rank test, we need to categorize this variable.

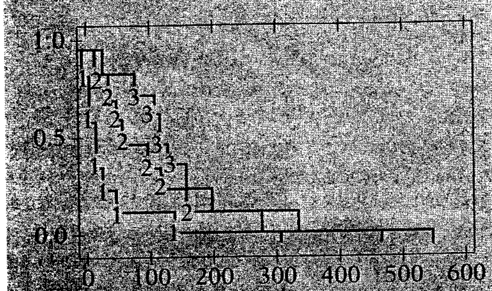
EXAMPLE (continued)

Performance Status Categories

Group #	Categories	Size
1	0–59	52
2	60–74	50
3	75–100	35

If, for the performance status variable, we choose the categories 0–59, 60–74, and 75–100, we obtain three groups of sizes 52, 50, and 35, respectively.

KM curves for performance status groups



The KM curves for each of three groups are shown here. Notice that these curves appear to be quite different. A test of significance of this difference is provided by the log–rank statistic.

Group	Size	% Cen	LQ	Med	UQ	0.95 Med CI	
1	52	3.846	11	20.5	61	18.000	29
2	50	6.000	51	93.5	228	54.000	118
3	35	11.429	103	133.0	201	112.000	182

Df: 2 Log–rank: 29.181, P-value: 0.0 Peto: 32.558, P-value: 0.0

A printout of descriptive information about the three KM curves together with the log–rank test and the Peto test results are shown here. These results were obtained using the SPIDA package.

$G = 3$ groups; $df = G - 1 = 2$

Log–rank test is highly significant.

Conclude: significant difference among three survival curves.

Because three groups are being compared here, $G = 3$ and the degrees of freedom for the log–rank test is thus $G - 1$, or 2. The log–rank statistic is computed to be 29.181, which has a P-value of zero to three decimal places. Thus, the conclusion from the log–rank test is that there is a highly significant difference among the three survival curves for the performance status groups.

Peto test also significant.

Note, also, that in this example, the Peto test is also highly significant.

VI. The Peto Test

An alternative to log–rank test

The Peto test was suggested as an alternative to the log–rank test by Prentice and Marek ("A Qualitative Discrepancy Between Censored Rank Tests," *Biometrics* 35: 861–867, 1979).

Log–rank uses:

$$O_i - E_i = \sum_j (m_{ij} - e_{ij}), \quad i = \text{group\#}, \; j = j\text{th failure time}$$

same weight = 1

In describing the difference between these two tests, recall that the log–rank test uses the summed observed minus expected score $O–E$ in each group to form the test statistic. This simple sum gives the same weight—namely, unity—to each failure time when combining observed minus expected failures in each group.

Peto test: $\text{weight} = n_j = \sum_{i=1}^{G} n_{ij}$

$$\text{Weighted average} = \frac{\sum_j n_j (m_{ij} - e_{ij})}{\sum_j n_j}$$

In contrast, the Peto test weights the observed minus expected score at time t_j by the number at risk, n_j, over all groups at time t_j. Thus, instead of a simple sum, the Peto test uses a weighted average of observed minus expected score, as shown here.

Computer performs calculations.

Peto statistic $\sim\chi^2$ with $G - 1$ *df*

The above formulae are not really important computationally, because a computer program can perform the calculations easily. The Peto test statistic, like the log–rank statistic, has approximately a large sample chi-square distribution with $G - 1$ degrees of freedom, where G is the number of survival curves being compared.

Peto test:
Emphasizes beginning of survival curve; early failures receive larger weights.

Nevertheless, the different formulae we have described indicate that the Peto test places more emphasis on the information at the beginning of the survival curve where the number at risk is large. Thus, early failures receive larger weights while failures in the tail of the survival curve receive smaller weights.

Log–rank test:
Emphasizes tail of the survival curve; equal weight given to each failure time.

In contrast, the log–rank test emphasizes failures in the tail of the survival curve, where the number at risk decreases over time, yet equal weight is given to each failure time.

Peto test is not necessarily a conservative test (when compared to log–rank test).

Despite the above differences between the log–rank and Peto tests, the Peto test is not necessarily a conservative test, because its numerical value may be either smaller or larger than the log–rank test, depending on the data being considered.

Choose:
1. Peto test if we want more weight given to earlier part of survival curve;
2. log–rank test, if otherwise.

In choosing between the log–rank test and the Peto test, we suggest using the Peto test if we want to give more weight to the earlier part to the survival curve where there are larger numbers at risk. Otherwise, choose the log–rank test. This choice of emphasizing earlier failure times may rest on clinical features of one's study. A discussion of the relative merits of these tests as well as some other alternatives is described by Harris and Albert in *Survivorship Analysis for Clinical Studies,* Marcel Dekker, 1991.

EXAMPLE: Remission Data

Results for Treatment Status:

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

chi2 log-rank: 16.793, P-value: 0.0, Peto: 9.954, P-value: 0.002

We illustrate the Peto test using examples shown earlier. For the remission data, a comparison of the treatment and placebo groups—with 21 subjects in each—yielded log–rank and Peto tests shown again here. Notice that both the log–rank and Peto tests are highly significant, although the Peto test yields a smaller chi-square value in this example.

EXAMPLE: vets.dat

Results for Performance Status (3 groups)

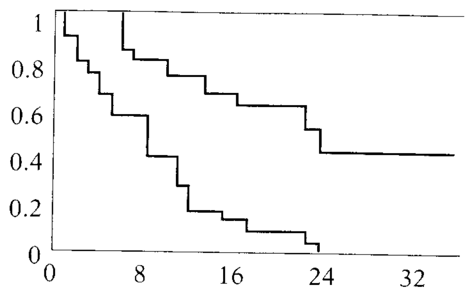
Group	Size	% Cen	LQ	Med	UQ	0.95 Med CI	
[illegible]	52	3.846	11	20.5	61	18.000	29
[illegible]	50	6.000	51	93.5	228	54.000	118
[illegible]	52	11.429	103	133.0	201	112.000	182

chi2 log-rank: 29.181, P-value: 0.0, Peto: 32.558, P-value: 0.0

As a second example, we consider the "vets.dat" data set previously described. The log–rank and Peto test results from comparing three groups of the variable performance status are again shown here. As in the above remission data example, both the log–rank and Peto tests are highly significant for the vets.dat data. Notice, however, that the Peto statistic of 32.558 is slightly higher than the log–rank statistic of 29.181.

VII. Summary

KM curves:



We now briefly summarize this presentation. First, we described how to estimate and graph survival curves using the Kaplan–Meier (KM) method.

$t_{(j)}$: jth ordered failure time

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

To compute KM curves, we must form a data layout that orders the failure times from smallest to largest. For each ordered failure time, the estimated survival probability is computed using the **product limit formula** shown here. Alternatively, this estimate can be computed as the product of the survival estimate for the previous failure time multiplied by the conditional probability of surviving past the current failure time.

Log–rank test:
H_0: common survival curve for all groups

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

log–rank statistic $\sim \chi^2$ with $G - 1$ df under H_0

G = # of groups

When survival curves are being compared, the log–rank test gives a statistical test of the null hypothesis of a common survival curve. For two groups, the log–rank statistic is based on the summed observed minus expected scores for a given group and its variance estimate. For several groups, a computer should always be used since the log–rank formula is more complicated mathematically. The test statistic is approximately chi-square in large samples with $G - 1$ degrees of freedom, where G denotes the number of groups being compared.

Peto test: use if more weight to earlier part of survival curve.

An alternative test is called the Peto test, which may be chosen if one wants to give more weight to the earlier part of the survival curves. This test is also a large sample chi-square test with $G - 1$ degrees of freedom.

Chapters

This presentation is now complete. You can review this presentation using the detailed outline that follows and then try the practice exercises and test.

Chapter 3 introduces the Cox proportional hazards (PH) model, which is the most popular mathematical modeling approach for estimating survival curves when considering several explanatory variables simultaneously.