

LOG-RANK TEST FOR TWO GROUPS

Prepared by Roberto Jimenez
Revised by Julia Soulakova

1. What research questions can be answered using the procedure?

Whenever you want to compare survival times to some event between two populations, you can use this test. For example, you want to assess whether male patients with a certain disease live longer than female do.

2. Data considerations

The log-rank test is based on several assumptions. These assumptions are:

- ✓ Sample is chosen randomly and independently from a larger population.
- ✓ Inclusion criteria are well stated, so that the definition of survival is consistent from subject (observation) to subject (observation).
- ✓ The entry criteria and baseline survival rate do not change over time
- ✓ The survival times of the censored subjects are the same, on average, as the survival of the remaining subjects.
- ✓ The variable designating group membership is independent of other patient covariates.
- ✓ There is no correlation among covariates

Notes on coding Variables:

Even though, the SAS procedures can deal with alphanumeric variables, for simplicity of data representation, categorical variables are coded using integers. For example, Sex can be code as 0 (male) and 1 (female). Similarly, the event indicator (status) is recorded as 1 (event happened) and 0 (event did not happen).

3. Main ideas and statistics behind the procedure

The main objective is to compare two survival curves by treatment group. Here we consider two groups. The hypotheses of interest are stated as follows:

$$H_0 : S_1(t) = S_2(t)$$

$$H_1 : S_1(t) \neq S_2(t),$$

where the null hypothesis states that the survival curves are identical in the two populations and the alternative states otherwise.

The Log-rank statistic is given by

$$\sum_{j=1}^r (d_{ij} - e_{ij}),$$

where $j=1,2,\dots,r$ denotes the distinct time points and $i=1,2$ denotes the group number. Then for each time interval, the observed number of events in each group is compared with the expected number of events, e_{ij} . If the null hypothesis is true, we expect $e_{ij} = n_{ij} \cdot d_j / n_j$, where

n_{ij} is the number at risk just prior to time j in group $i=1,2$.

n_j is the total number of cases that are at risk just prior to time j .

d_j is the total number of events at time j in both groups.

Then based on these statistics, the chi-square test statistic is calculated and the corresponding p-value is obtained.

4. Example (A small data set is used for simplicity of illustration).

The next data set corresponds to a survey administered to assess whether or not breast feeding moms are more likely to retain from smoking.

The variables are:

- ✓ Days: the number of days from delivery to smoking relapse or to the end of study
- ✓ Status: 1- woman started smoking, 0-otherwise
- ✓ Breastfeed: 1- woman breast feeds, 0-otherwise

The SAS code to run the procedure.

```

data momquit;
input days status Breastfeed;
datalines;
10      1      0
111     0      1
100     1      1
23      1      1
232     0      0
312     0      1
45      1      1
54      1      1
67      1      0
67      1      0
100     0      1
100     1      1
run;

proc lifetest data=momquit plots=(s) graphics;
time days*status(0);
strata Breastfeed;
symbol1 v=none color=black line=1;
symbol2 v=none color=black line=2;
run;

```

The SAS output

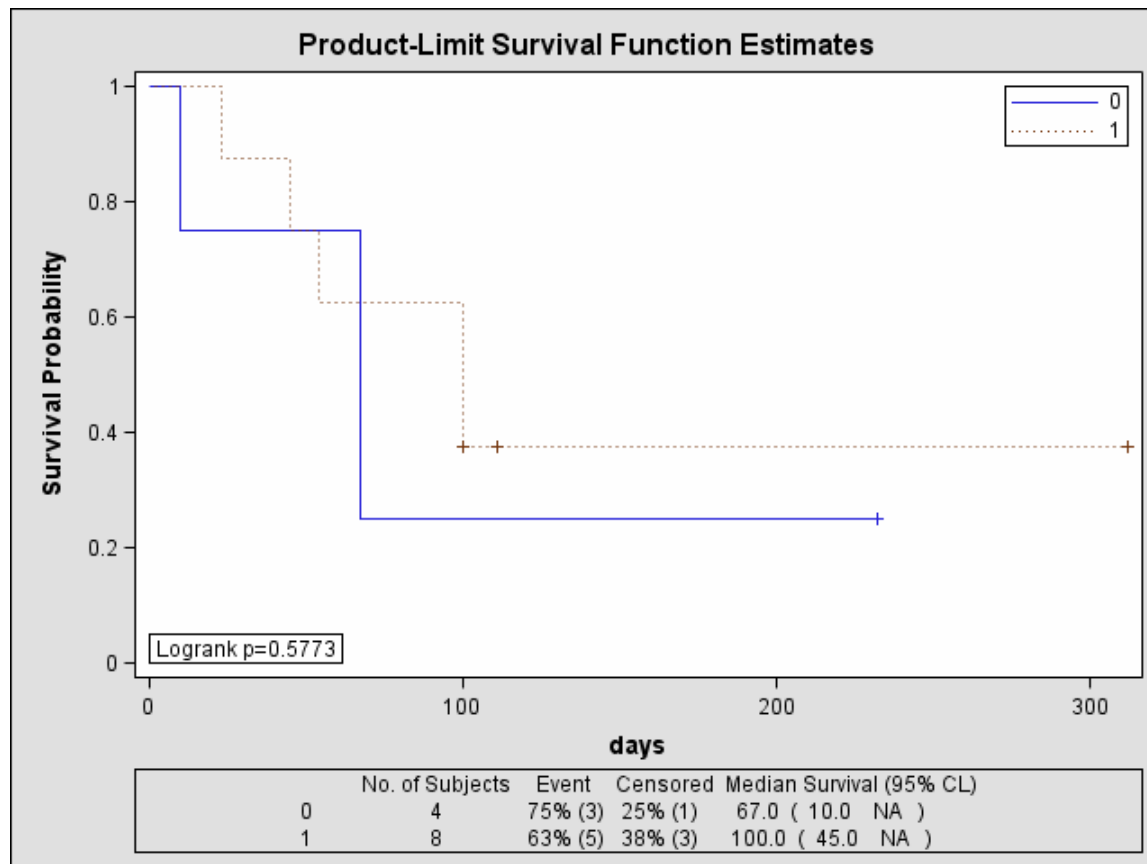
Rank Statistics		
Breastfeed	Log-Rank	Wilcoxon
0	0.67727	7.0000
1	-0.67727	-7.0000

Covariance Matrix for the Log-Rank Statistics		
Breastfeed	0	1
0	1.47680	-1.47680
1	-1.47680	1.47680

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.3106	1	0.5773
Wilcoxon	0.3807	1	0.5372
-2Log(LR)	0.1629	1	0.6865

From the output we obtain the log-rank statistics, 0.67727, squaring that number we have 0.45869. Then dividing by the estimated variance, 1.47680, we obtain our observed chi-square statistic, 0.3106. Next, we compare it with a 5% critical value corresponding to a chi-square distribution with 1 degree of freedom (3.84). Since $0.3106 < 3.84$, we fail to reject the null hypothesis. That is, there is no significant difference among the breast feeding and not breast feeding women in terms of the time to smoking relapse.

Graphs below illustrate the survival probabilities for breast feeding and not breast feeding women. Survival curves look similarly, which support our conclusion.



References:

1. Intuitive Biostatistics, by Harvey Motulsky. Oxford University Press Inc. 1995
2. Survival Analysis Using SAS: A Practical Guide By Paul D. Allison, SAS Press 1995.