

Biostatistics (MATH11230)

Comparing survival curves

Vanda Inácio

University of Edinburgh



Semester 1, 2022/2023

Comparing survival curves

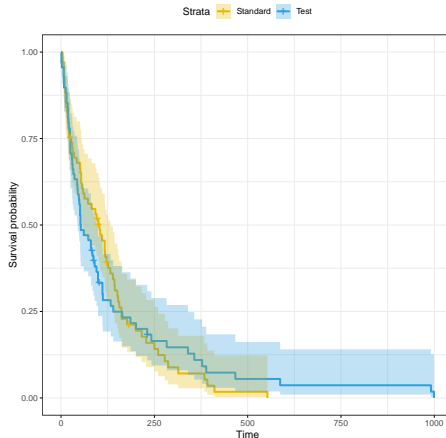
General context

- ↪ We have learned about point and interval estimation of survival curves.
- ↪ Often, there is additional information about the patients/individuals, and we would like to add this to our analysis.
- ↪ A very first step towards this end is to compare the survival curves from two groups.
- ↪ We have looked at the veteran dataset, available in the `survival` package.
- ↪ This was a study conducted by the US Veterans Administration. Male patients with advanced inoperable lung cancer were given either a standard therapy or a test chemotherapy.

Comparing survival curves

General context

- Kaplan–Meier estimates, along with 95% confidence bands, of the survival function in the two treatment groups.



Comparing survival curves

General context

- ↪ We can see that the two estimated Kaplan–Meier curves slightly differ and that there is a good amount of overlap between the 95% pointwise confidence bands.
- ↪ Clearly, however, it would also be nice to formally test for a difference between the two groups survival curves.
- ↪ Specifically, we want to test the null hypothesis $H_0 : S_1 = S_2$, where S_1 and S_2 are the survival functions for two different groups of patients (for the veteran dataset example, S_1 and S_2 are the survival curves for the two treatment groups).

Comparing survival curves

General context

- ↪ Of course, one possibility would be to compute the survival curves for each group and compare the proportions surviving at any specific time.
- ↪ The weakness of this approach is that it does not provide a comparison of the total survival experience of the two groups, but rather gives a comparison at some arbitrary time point(s).
- ↪ If we knew that S_1 and S_2 were in the same parametric family (e.g., $S_j(t) = e^{-\lambda_j t}$, $j = 1, 2$), then the null hypothesis would reduce to testing whether the parameters were the same (e.g., $\lambda_1 = \lambda_2$).
- ↪ However, we instead want a nonparametric test, that is, a test whose validity does not depend on any parametric assumptions.
- ↪ The log rank test is the most widely used approach to compare survival curves. It adapts methods for analysing 2×2 contingency tables (Mantel, 1966) to accommodate censoring.
- ↪ Here we introduce the log rank test from a heuristic perspective and to discuss popular extensions.

Comparing survival curves

Log rank test

- ↪ We will use the same notation that we have used for the Kaplan–Meier estimator, with straightforward extensions to accommodate multiple groups.
- ↪ Let $0 = t_0 < t_1 < t_2 < \dots < t_J < t_{J+1} = \infty$ denote the pooled event times (times at which any subject in either group had an event).
- ↪ Let d_{1j} denote the number of events at time t_j in group 1, d_{2j} the number of events at time t_j in group 2 and d_j denote the total number of events at time t_j .
- ↪ Similarly, let n_{1j} be the number of individuals at risk in group 1 at time t_j , n_{2j} be the number of individuals at risk in group 2 at time t_j and n_j denote the total number of individuals at risk at time t_j .

Comparing survival curves

Log rank test

- ↪ The information available at each event time $t_j, j = 1, \dots, J$, is usually represented in a 2×2 contingency table

	# with an event at t_j	# without an event at t_j	Total
Group 1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
Group 2	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Total	d_j	$n_j - d_j$	n_j

- ↪ Formally, we are testing

$$\begin{aligned} H_0 : S_1(t) &= S_2(t), & \text{for all } t \leq t_J, \\ H_1 : S_1(t) &\neq S_2(t), & \text{for some } t \leq t_J. \end{aligned}$$

- ↪ The idea behind the log rank test is to compare, at each event time, the observed number of events with the expected number of events under the null hypothesis of no association between events and grouping.

Comparing survival curves

Log rank test

- ↪ Conditional on the four marginal totals in the 2×2 contingency table, in the previous slide, being fixed, a single element (say d_{1j}) defines the table.
- ↪ One can show that, conditional on the marginal totals and under the null hypothesis that survival is independent of the group being true, d_{1j} follows an hypergeometric distribution with mean

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

which therefore represents, under H_0 , the expected number of events in group 1 at time t_j .

- ↪ This value is intuitively appealing since under the null hypothesis that the probability of an event occurring at time t_j does not depend on the group that an individual is in, the probability of an event at t_j is d_j/n_j . Multiplying this by n_{1j} gives e_{1j} as the expected number of events in group 1 at time t_j .

Comparing survival curves

Log rank test

- ↪ The idea is then to compare the number of observed, d_{1j} , and expected, e_{1j} , events and to combine the information from all event times.
- ↪ The most straightforward way of doing this is to sum the differences $d_{1j} - e_{1j}$ over the total number of event times, J , in the two groups.
- ↪ The resulting test statistic is given by

$$U_L = \sum_{j=1}^J (d_{1j} - e_{1j}) = \sum_{j=1}^J \left(d_{1j} - n_{1j} \frac{d_j}{n_j} \right).$$

- ↪ Obviously, this is equivalent to $\sum_{j=1}^J d_{1j} - \sum_{j=1}^J e_{1j}$, which is the difference between the observed and expected number of events in group 1, sometimes abbreviated as $O - E$ (for observed minus expected).

Comparing survival curves

Log rank test

- ↪ This U_L statistic has zero mean since $E(d_{1j} | H_0) = e_{1j}$.
- ↪ Moreover, since the event times are independent of one another, the variance of U_L is simply the sums of the variances of the d_{1j} s

$$\text{var}(U_L) = \text{var} \left(\sum_{j=1}^J d_{1j} - e_{1j} \right) = \sum_{j=1}^J \text{var}(d_{1j}).$$

- ↪ Because d_{1j} follows a hypergeometric distribution, under H_0 , the variance of d_{1j} is given by

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

Comparing survival curves

Log rank test

↪ One can show that U_L has an approximate normal distribution, when the number of events is not too small.

↪ It then follows that

$$\frac{U_L}{\sqrt{\text{var}(U_L)}} \sim N(0, 1).$$

↪ We thus have that

$$\frac{U_L^2}{\text{var}(U_L)} \sim \chi_1^2.$$

↪ The statistic $U_L^2/\text{var}(U_L)$ summarises the extent to which the observed event times in the two groups deviate from those expected under the null hypothesis of no group differences.

↪ The larger the value of this statistic, the greater the evidence against the null hypothesis.

Comparing survival curves

Log rank test

- ↪ The log rank test is sometimes also called the Mantel–Haenszel test, the Mantel–Cox test or the Peto–Mantel–Haenszel test.
- ↪ The log rank test is particularly powerful when the ratio between the two hazard functions being compared is constant across time, i.e.,

$$\frac{h_2(t)}{h_1(t)} = \text{constant}.$$

Comparing survival curves

Log rank test

→ Let us consider the following toy example:

Group 1: 3.1, 6.8⁺, 9, 9, 11.3⁺, 16.2

Group 2: 8.7, 9, 10.1⁺, 12.1⁺, 18.7, 23.1⁺

→ We have that $J = 5$ and the distinct event times are: 3.1, 8.7, 9, 16.2, and 18.7.

→ We have that for $t_1 = 3.1$

	# with an event at t_1	# without an event at t_1	Total
Group 1	1	5	6
Group 2	0	6	6

→ We have that for $t_2 = 8.7$

	# with an event at t_2	# without an event at t_2	Total
Group 1	0	4	4
Group 2	1	5	6

Comparing survival curves

Log rank test

↪ We have that for $t_3 = 9$

	# with an event at t_3	# without an event at t_3	Total
Group 1	2	2	4
Group 2	1	4	5

↪ We have that for $t_4 = 16.2$

	# with an event at t_4	# without an event at t_4	Total
Group 1	1	0	1
Group 2	0	2	2

↪ We have that for $t_5 = 18.7$

	# with an event at t_5	# without an event at t_5	Total
Group 1	0	0	0
Group 2	1	1	2

Comparing survival curves

Log rank test

↪ We have that $\sum_{j=1}^5 d_{1j} = 1 + 0 + 2 + 1 + 0 = 4$.

↪ We have that

$$\sum_{j=1}^5 e_{1j} = 6 \times \frac{1}{12} + 4 \times \frac{1}{10} + 4 \times \frac{3}{9} + 1 \times \frac{1}{3} + 0 = 2.566667.$$

↪ After doing the calculation, one has that $\text{var}(U_L) = \sum_{j=1}^5 v_{1j} = 1.267778$.

↪ Finally, we have that the value of the test statistic is $(4 - 2.566667)^2 / 1.267778 = 1.620508$.

↪ The resulting p-value is 0.2 and so there is no evidence to reject the null hypothesis that the survival curves of the two groups are the same.

Comparing survival curves

Log rank test

- ↪ We have illustrated the calculations involved in the log rank test for this very toy example. Of course, for larger (not even much larger!!) examples, these calculations are done in R.
- ↪ The function `survdif` in the `survival` package implements the log rank test.
- ↪ For instance, for the ‘motivating’ veteran dataset, there are 101 distinct event times across the two treatment groups!
- ↪ In this case, the p-value is 0.9, and not surprisingly, we do not reject the null hypothesis of no difference in the survival of the two treatment groups.

Comparing survival curves

Weighted log rank test

↪ The U_L statistic of the log rank test

$$U_L = \sum_{j=1}^J (d_{1j} - e_{1j})$$

gives the same weight to the difference between the observed and expected number of events at all timepoints.

↪ One natural extension is the family of weighted log rank tests, whose test statistic is

$$U = \sum_{j=1}^J w_j (d_{1j} - e_{1j}),$$

and where the weights $\{w_j\}_{j=1}^J$ are chosen to emphasise or deemphasise various time points.

↪ The variance of this weighted sum is obtained as before, $\text{var}(U) = \sum_{j=1}^J w_j^2 v_{1j}$, and we still have that $U^2 / \text{var}(U) \sim \chi_1^2$.

Comparing survival curves

Weighted log rank test

- ↪ For example, one reasonable weighting scheme would be to weight time points by the number of individuals at risk at time t_j , that is,

$$w_j = n_j, \quad j = 1, \dots, J.$$

- ↪ The idea is to give less weight to the difference between observed and expected number of events for the event times where there are less information, i.e, when the risk set becomes small.
- ↪ It will therefore give more emphasis on the deviation of the null hypothesis earlier in time when there are more individuals still at risk, and less emphasis to the deviation from the null hypothesis of equality of the two curves occurring later in the tails of the survival functions (i.e., for largest event times).
- ↪ This is known as the Gehan, Gehan–Breslow, or Gehan–Wilcoxon test.

Comparing survival curves

Weighted log rank test

- ↪ A somewhat similar idea is to consider $w_j = \hat{S}(t_j)$, with $\hat{S}(\cdot)$ an estimate of the common survival function of both groups.
- ↪ This is known as the Peto–Prentice or Peto–Peto test and also gives more weight to early differences compared to late differences, but independently of the number individuals still at risk.
- ↪ Compared to the log rank test, both tests place more emphasis on earlier differences.
- ↪ The `survdif` function has an option, `rho`, that can be set to 1 in order to perform the Peto–Prentice test.
- ↪ The package `survminer` also allows to perform both weighted tests.

Comparing survival curves

Extension to multiple groups

- ↪ The log rank test can be extended to compare three or more groups of event times data.
- ↪ Suppose that the survival distributions of G groups of event times data are to be compared, for $G \geq 2$.
- ↪ We then define the analogous of the U-statistics for comparing the observed and expected number of events in groups $1, 2, \dots, G - 1$.

↪ Let

$$U_{Lg} = \sum_{j=1}^J \left(d_{gj} - n_{gj} \frac{d_j}{n_j} \right) = \sum_{j=1}^J (d_{gj} - e_{gj}), \quad g = 1, \dots, G - 1.$$

- ↪ These quantities are then expressed in the form of a vector with $G - 1$ components

$$\mathbf{U}_L = (U_{L1}, \dots, U_{L,G-1})'.$$

Comparing survival curves

Extension to multiple groups

- ↪ We also need the expressions for the variances of the U_{Lg} and for the covariances between pairs of values.
- ↪ In particular, the covariance between U_{Lg} and $U_{Lg'}$ is given by

$$V_{Lgg'} = \sum_{j=1}^J \frac{n_{gj}d_j(n_j - d_j)}{n_j(n_j - 1)} \left(\delta_{gg'} - \frac{n_{g'j}}{n_j} \right),$$

for $g, g' = 1, 2, \dots, G - 1$ and where $\delta_{gg'} = 1$ if $g = g'$ and 0 otherwise.

- ↪ These terms are then assembled in the form of a variance-covariance matrix, \mathbf{V}_L .
- ↪ For example, in the comparison of three groups of event times data, this matrix is

$$\mathbf{V}_L = \begin{pmatrix} V_{L11} & V_{L12} \\ V_{L12} & V_{L22} \end{pmatrix}.$$

Comparing survival curves

Extension to multiple groups

- ↪ Finally, in order to test the null hypothesis of no group differences, we make use of the following result

$$\mathbf{U}'_L \mathbf{V}_L^{-1} \mathbf{U}_L \sim \chi^2_{G-1}.$$

- ↪ Note that in this setup, we have G groups but only $G - 1$ independent counts in the contingency table because we are conditioning on the margins, and therefore only $G - 1$ degrees of freedom.
- ↪ The function `survdif` also implements this case.