

Confidentiality for Cox model

The WebDISCO method for estimating a survival model does not ensure data confidentiality because predictors can be retrieved by the central server. This occurs due to sending data at every event time. When only one event happens during a given period, only the predictors of that individual vary, making it possible to identify them.

To address this issue, it is important to ensure that there is never a single event occurring at a single event time. Therefore, data can be divided into intervals where many patients are grouped. Various methods are available for this purpose.

Averaging

The data is ordered by time, and values are grouped and given a new time, which is the average of all values of time.

Example.

For the example, we assume we want to group data into groups of 2.

| time | p_1 | p_2 |
|------|-------|-------|
| 2 | 43 | 0 |
| 4 | 24 | 0 |
| 5 | 41 | 1 |
| 6 | 37 | 1 |
| 9 | 53 | 0 |
| 11 | 33 | 1 |
| 12 | 39 | 1 |
| 17 | 45 | 0 |

| time | p_1 | p_2 |
|------|-------|-------|
| 3 | 43 | 0 |
| 3 | 24 | 0 |
| 5.5 | 41 | 1 |
| 5.5 | 37 | 1 |
| 10 | 53 | 0 |
| 10 | 33 | 1 |
| 14.5 | 39 | 1 |
| 14.5 | 45 | 0 |

Where p_1 and p_2 are predictors.

Uniform Intervals

The period of the study is split into uniform intervals that guarantee at least x subjects per event time.

Example

To have at least 2 values in every interval, the intervals must be bigger than the biggest difference between values. Here, the biggest difference is $17 - 12 + 1 = 6$. The intervals are of size 6.

| time | p_1 | p_2 |
|------|-------|-------|
| 2 | 43 | 0 |
| 4 | 24 | 0 |
| 5 | 41 | 1 |
| 6 | 37 | 1 |
| 9 | 53 | 0 |
| 11 | 33 | 1 |
| 12 | 39 | 1 |
| 17 | 45 | 0 |

| time | p_1 | p_2 |
|------|-------|-------|
| 4.5 | 43 | 0 |
| 4.5 | 24 | 0 |
| 4.5 | 41 | 1 |
| 4.5 | 37 | 1 |
| 10.5 | 53 | 0 |
| 10.5 | 33 | 1 |
| 10.5 | 39 | 1 |
| 16.5 | 45 | 0 |

Intervals: 2 – 7, 8 – 13, 14 – 19.

The last value, 17, is not in an interval of adequate size. To solve that issue, the last two intervals are merged.

Non-uniform Intervals

The period of the study is split into non-uniform intervals. These intervals are the smallest size possible.

Example

Here, the goal is to form the smallest intervals.

| time | p_1 | p_2 |
|------|-------|-------|
| 2 | 43 | 0 |
| 4 | 24 | 0 |
| 5 | 41 | 1 |
| 6 | 37 | 1 |
| 9 | 53 | 0 |
| 11 | 33 | 1 |
| 12 | 39 | 1 |
| 17 | 45 | 0 |

| time | p_1 | p_2 |
|------|-------|-------|
| 3 | 43 | 0 |
| 3 | 24 | 0 |
| 5.5 | 41 | 1 |
| 5.5 | 37 | 1 |
| 10 | 53 | 0 |
| 10 | 33 | 1 |
| 14.5 | 39 | 1 |
| 14.5 | 45 | 0 |

Intervals: 2 – 4, 5 – 6, 9 – 11, 12 – 17.

However, these intervals must be the same for all sites.