STATISTICAL ANALYSIS OF SURVIVAL DATA IN CLINICAL RESEARCH

In this course:

Presentation of statistical methods to describe mortality and prognosis of patients suffering a particular disease in a given period of time.

Also some discussion of related methods used in epidemiological studies.

Basic problem: The study of occurrence of health-related events in time.

More generally: Several events per individual.

Examples:

- Duration of unemployment
- Life insurance
- Product reliability
- Mortality
- Chronic disease incidence

More generally: Several events per individual.

Examples:

- Cause-specific mortality
- Recurrent diseases
- Repeated events (recurrence)

Some standard situations:

a) "Alive" Disease -> "Alive" Disease
b) "Alive" Disease -> "Dead" Disease -> "Alive" Disease
c) "Alive" Disease -> "Dead" Disease -> "Married" -> "Never Married"
d) "Dead" Disease -> "Dead" Disease

In this course: Presentation of statistical methods used in particular disease in a given period of time.
SURVIVAL DATA

Two aspects:
A waiting time
Occurrence of an event

A continuous and a discrete part.

Choice of time scale:
- Time since entry
- Birth, i.e., time = age
- Time since diagnosis

Additional feature, which complicates the data:
that the waiting time exceeds a specified value.

(right) censoring
end of the period of observation.

The event has not (yet) occurred at the
occurrence of the event.

The period of observation is terminated by
(right)
censoring.

If the period of observation is

(1, 0)


data for each individual in the study:

DATE: For each individual in the study:

two aspects:

Occurrence of an event
A waiting time

Points:
- A waiting time, but what is the appropriate starting

Time of scale:
- A continuous and a discrete part.

Risk
Prospects

How do we quantify:
Describing occurrence in time of a health-related

or something else?

Probabilities are an obvious choice, but not the only

Possibility:

Is it dangerous to travel by airplane?
DESCRIBING THE PROGNOSIS, HOW?

Survival function = Survivorship function

\[ S(t) = \text{The probability of being alive (event-free) at time } t \]

If all patients are followed until death, i.e. data with no censored observations:

The survival function is estimated by

\[ \frac{\text{number of patients alive at time } t}{\text{number of patients alive at time } 0} \]

Presence of censored observations: The simple proportion is no longer appropriate. Instead, the survival function is estimated by

The life table method or Kaplan-Meier’s method

ALTERNATIVE DESCRIPTION OF MORTALITY/SURVIVAL

In Demography and Insurance Mathematics:

Mortality rate = Death intensity = Force of Mortality = Hazards rate

In Actuarial Science:

Survival function = Survival curve

DESCRIPTING THE PROGNOSIS, HOW?
Estimating the mortality rate:

Divide the follow-up period up in small intervals:

\[
\text{Closing Date} = 0
\]

For each small time interval compute:

- Number of deaths in the interval
- Total time at risk in the interval

The ratio

\[
\frac{\text{Number of deaths in the interval}}{\text{Total time at risk in the interval}} = (i)\gamma
\]

is an estimate of the mortality rate at this point in time.

\[\text{Note:} \]

A probability, i.e. \(0 \leq S(t) \leq 1\)

No dimension

\[m(t) = \gamma \] Dimension = per time unit

A rate, i.e. \(0 \leq m(t) \leq 1\)

Dimension = per time unit

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]

\[m(t) = \gamma \exp(-\int_0^t \gamma \, ds)\]

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]

The mortality rate is the relative rate of change of the survival probability or the relative rate of change in the number of survivors.

\[\frac{d}{dt} S(t) = -\gamma S(t)\]

\[\text{Integrating hazard} = \int_0^t \gamma \, ds = \text{Integrating mortality rate or mortality rate, Death intensity} = (i)\gamma\]

Survival function

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]

\[\text{Estimating the mortality rate:} \]

Divide the follow-up period up in small intervals:

Date Closing

Time

For each small time interval compute:

- Number of deaths in the interval
- Total time at risk in the interval

The ratio

\[\frac{\text{Number of deaths in the interval}}{\text{Total time at risk in the interval}} = (i)\gamma\]

\[\text{Note:} \]

A probability, i.e. \(0 \leq S(t) \leq 1\)

No dimension

\[m(t) = \gamma \] Dimension = per time unit

A rate, i.e. \(0 \leq m(t) \leq 1\)

Dimension = per time unit

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]

\[m(t) = \gamma \exp(-\int_0^t \gamma \, ds)\]

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]

\[\text{Integrating hazard} = \int_0^t \gamma \, ds = \text{Integrating mortality rate or mortality rate, Death intensity} = (i)\gamma\]

Survival function

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]

\[\text{Estimating the mortality rate:} \]

Divide the follow-up period up in small intervals:

Date Closing

Time

For each small time interval compute:

- Number of deaths in the interval
- Total time at risk in the interval

The ratio

\[\frac{\text{Number of deaths in the interval}}{\text{Total time at risk in the interval}} = (i)\gamma\]

\[\text{Note:} \]

A probability, i.e. \(0 \leq S(t) \leq 1\)

No dimension

\[m(t) = \gamma \] Dimension = per time unit

A rate, i.e. \(0 \leq m(t) \leq 1\)

Dimension = per time unit

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]

\[m(t) = \gamma \exp(-\int_0^t \gamma \, ds)\]

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]

\[\text{Integrating hazard} = \int_0^t \gamma \, ds = \text{Integrating mortality rate or mortality rate, Death intensity} = (i)\gamma\]

Survival function

\[S(t) = \exp(-\int_0^t \gamma \, ds)\]
ESTIMATION OF THE SURVIVAL FUNCTION

WITH CENSORED DATA

Standard survival analysis:

- event: death from all causes
- For each individual:
  - Does he/she experience the event?
  - When?
  - Event-free for how long?
- If data contain no censored observations then
  \[ Y(t) = n \times \text{empirical survival function} \]
  \[ N(t) = n \times \text{empirical distribution function} \]
- With censored data these relations are no longer true.

A simple example

Data:

<table>
<thead>
<tr>
<th>Survival time in days measured from time of entry</th>
<th>Censored observations marked by +</th>
</tr>
</thead>
<tbody>
<tr>
<td>55 138 168 202+ 220+</td>
<td>Number of events N(t)</td>
</tr>
<tr>
<td>100 161 214+ 222+ 238</td>
<td>Number at risk Y(t)</td>
</tr>
</tbody>
</table>

Sample size = \( n = 12 \)

\[ Y(t) = n \times \text{empirical survival function} \]
\[ N(t) = n \times \text{empirical distribution function} \]
KAPLAN-MEIER'S METHOD

The calculations involves the following steps:

1) The observations are sorted in ascending order.

Convention for ties: Uncensored observations are placed before the censored observations.

<table>
<thead>
<tr>
<th>Time</th>
<th>At Risk</th>
<th>Survival</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>12</td>
<td>0.9000</td>
<td>1</td>
</tr>
<tr>
<td>61+</td>
<td>11</td>
<td>0.8367</td>
<td>0.9000</td>
</tr>
<tr>
<td>74</td>
<td>10</td>
<td>0.8489</td>
<td>0.8367</td>
</tr>
<tr>
<td>81</td>
<td>9</td>
<td>0.8111</td>
<td>0.8489</td>
</tr>
<tr>
<td>93+</td>
<td>8</td>
<td>0.7333</td>
<td>0.8111</td>
</tr>
<tr>
<td>122+</td>
<td>7</td>
<td>0.6111</td>
<td>0.7333</td>
</tr>
<tr>
<td>138</td>
<td>6</td>
<td>0.4890</td>
<td>0.6111</td>
</tr>
<tr>
<td>141</td>
<td>5</td>
<td>0.4098</td>
<td>0.4890</td>
</tr>
<tr>
<td>197</td>
<td>4</td>
<td>0.3617</td>
<td>0.4098</td>
</tr>
<tr>
<td>220+</td>
<td>3</td>
<td>0.3333</td>
<td>0.3617</td>
</tr>
<tr>
<td>224</td>
<td>2</td>
<td>0.2222</td>
<td>0.3333</td>
</tr>
</tbody>
</table>

2) For each time of death i compute

\[ S(t)_i = \prod_{j=1}^{i} \left( 1 - \frac{d_j}{n_j} \right) \]

Then Kaplan-Meier's estimate (also denote the product-limit estimate)

\[ \hat{S}(t)_i = \prod_{j=1}^{i} \left( 1 - \frac{d_j}{n_j} \right) \]

The example:

\[ \hat{S}(55) = 0.9000 \]
\[ \hat{S}(74) = 0.8489 \]
\[ \hat{S}(81) = 0.8111 \]
\[ \hat{S}(93) = 0.7333 \]
\[ \hat{S}(122) = 0.6111 \]
\[ \hat{S}(138) = 0.4890 \]
\[ \hat{S}(141) = 0.4098 \]
\[ \hat{S}(197) = 0.3617 \]
\[ \hat{S}(220) = 0.3333 \]
\[ \hat{S}(224) = 0.2222 \]
\[ \hat{S}(220+3) = 0.1667 \]

The example involves the following steps:

1) The observations are sorted in ascending order.
2) For each time of death i compute \( \hat{S}(t)_i = \prod_{j=1}^{i} \left( 1 - \frac{d_j}{n_j} \right) \)

The calculations involves the following steps:

1) The observations are sorted in ascending order.
2) For each time of death i compute \( \hat{S}(t)_i = \prod_{j=1}^{i} \left( 1 - \frac{d_j}{n_j} \right) \)

The example:

\[ \hat{S}(55) = 0.9000 \]
\[ \hat{S}(74) = 0.8489 \]
\[ \hat{S}(81) = 0.8111 \]
\[ \hat{S}(93) = 0.7333 \]
\[ \hat{S}(122) = 0.6111 \]
\[ \hat{S}(138) = 0.4890 \]
\[ \hat{S}(141) = 0.4098 \]
\[ \hat{S}(197) = 0.3617 \]
\[ \hat{S}(220) = 0.3333 \]
\[ \hat{S}(224) = 0.2222 \]

The example involves the following steps:

1) The observations are sorted in ascending order.
2) For each time of death i compute \( \hat{S}(t)_i = \prod_{j=1}^{i} \left( 1 - \frac{d_j}{n_j} \right) \)

The example:

\[ \hat{S}(55) = 0.9000 \]
\[ \hat{S}(74) = 0.8489 \]
\[ \hat{S}(81) = 0.8111 \]
\[ \hat{S}(93) = 0.7333 \]
\[ \hat{S}(122) = 0.6111 \]
\[ \hat{S}(138) = 0.4890 \]
\[ \hat{S}(141) = 0.4098 \]
\[ \hat{S}(197) = 0.3617 \]
\[ \hat{S}(220) = 0.3333 \]
\[ \hat{S}(224) = 0.2222 \]

The example involves the following steps:

1) The observations are sorted in ascending order.
2) For each time of death i compute \( \hat{S}(t)_i = \prod_{j=1}^{i} \left( 1 - \frac{d_j}{n_j} \right) \)

The example:

\[ \hat{S}(55) = 0.9000 \]
\[ \hat{S}(74) = 0.8489 \]
\[ \hat{S}(81) = 0.8111 \]
\[ \hat{S}(93) = 0.7333 \]
\[ \hat{S}(122) = 0.6111 \]
\[ \hat{S}(138) = 0.4890 \]
\[ \hat{S}(141) = 0.4098 \]
\[ \hat{S}(197) = 0.3617 \]
\[ \hat{S}(220) = 0.3333 \]
\[ \hat{S}(224) = 0.2222 \]

The example involves the following steps:

1) The observations are sorted in ascending order.
2) For each time of death i compute \( \hat{S}(t)_i = \prod_{j=1}^{i} \left( 1 - \frac{d_j}{n_j} \right) \)
KAPLAN-MEIER'S ESTIMATE

Data:
A sample of size \( n \).
The number of observed deaths \( d \) (\( d \leq n \)).

In the example \( n = 12 \) and \( d = 7 \).

Ordered survival times (uncensored observations).

Kaplan-Meier's estimate:

\[ \hat{S}(t) = 1 - \text{empirical distribution function} \]

No censored observations:

Deaths occur before censoring

If censored and uncensored observations are tied:

\[ \hat{S}(t) = \prod_{i=1}^{m} \left( \frac{i-1}{m} \right) \]

where \( m = \) number of deaths at time \( t \).

If ties are present in the data:

\[ \hat{S}(t) = \prod_{i=1}^{m} \left( \frac{i-1}{m} \right) \]

The formula above assumes that all death times are distinct (no ties among uncensored observations).

A product: Each time of death contributes a factor to this product.

Entering the data:

1. Enter the data in Stata's spreadsheet (convenient only for small data sets) and name the variables, e.g.,

<table>
<thead>
<tr>
<th>Time</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.8</td>
<td>1</td>
</tr>
<tr>
<td>2.2</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>0</td>
</tr>
<tr>
<td>1.68</td>
<td>1</td>
</tr>
<tr>
<td>1.51</td>
<td>1</td>
</tr>
<tr>
<td>1.38</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1.22</td>
<td>0</td>
</tr>
<tr>
<td>0.93</td>
<td>1</td>
</tr>
<tr>
<td>1.18</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.61</td>
<td>1</td>
</tr>
<tr>
<td>0.55</td>
<td>1</td>
</tr>
</tbody>
</table>

Time and status (use time, failurestatus=?)

USING STATA FOR THE ANALYSIS

1) Enter the data in Stata's spreadsheet (convenient only for small data sets) and name the variables, e.g.,

<table>
<thead>
<tr>
<th>Time</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.8</td>
<td>1</td>
</tr>
<tr>
<td>2.2</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>0</td>
</tr>
<tr>
<td>1.68</td>
<td>1</td>
</tr>
<tr>
<td>1.51</td>
<td>1</td>
</tr>
<tr>
<td>1.38</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1.22</td>
<td>0</td>
</tr>
<tr>
<td>0.93</td>
<td>1</td>
</tr>
<tr>
<td>1.18</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.61</td>
<td>1</td>
</tr>
<tr>
<td>0.55</td>
<td>1</td>
</tr>
</tbody>
</table>

2) Define the data to be survival time data

a) On the command line write:

or use the mouse

Statistics - Survival analysis – Setup & Utilities
- Declare data to be survival time data
- Specify Time variable (time), Failure variable (status) and failure value (1) in the menu.
3) To get a Kaplan-Meier estimate as tabular output, write on the command line:

```
sts graph, censored
```

Some additional options (to be placed after the comma) include:

- `graph`: Generate a graph with the Kaplan-Meier estimate
- `ci`: Include 95% pointwise confidence limits on the graph.

To get a Kaplan-Meier survival estimate with 95% confidence limits:

```
sts graph, censored
```

4) To get a graph as output, write on the command line:

```
sts graph
```

Some additional options (to be placed after the comma) include:

- `ci`: Include 95% pointwise confidence limits on the graph.
- `overlay`: Overlay the survival estimates from different groups.
- `xline`: Draw a vertical line at a specified time point.
- `yline`: Draw a horizontal line at a specified survival probability.

The output will be displayed in a graph format, showing the Kaplan-Meier survival function over time.
AN ESTIMATE OF THE INTEGRATED HAZARD: NELSON-AALEN'S METHOD

The calculations involve the following steps:

1) The observations are sorted in ascending order.

2) For each time of death \(t\), compute

\[
\hat{Y}(t) = \sum_{i=1}^{n} \frac{1}{\Delta t_i}
\]

where \(\Delta t_i = t_i - t_{i-1}\) is the interval between death times.

3) Nelson-Aalen's estimate

\[
\hat{L}_t = \sum_{i=1}^{n} \hat{Y}(t)
\]

The example:

\[
\begin{align*}
\hat{L} &= 0.083 \\
\hat{L} &= \hat{L} + 0.183 \\
\hat{L} &= \hat{L} + 0.294 \\
\hat{L} &= \hat{L} + 0.461 \\
\hat{L} &= \hat{L} + 0.661 \\
\hat{L} &= \hat{L} + 0.911 \\
\hat{L} &= \hat{L} + 1.291 \\
\hat{L} &= \hat{L} + 1.911
\end{align*}
\]

NELSON-AALEN'S ESTIMATE

Data and notation: As for the Kaplan-Meier estimate.

1) For each time of death, compute \(\hat{Y}(t) = \frac{1}{\Delta t_i}\) and \(\hat{L}_t = \sum_{i=1}^{n} \hat{Y}(t)\).

2) The observations are sorted in ascending order.

3) Nelson-Aalen's estimate

\[
\hat{Y}(t) = \frac{1}{\Delta t_i}
\]

\[
\hat{L}_t = \sum_{i=1}^{n} \hat{Y}(t)
\]
USING STATA FOR THE ANALYSIS

Each data set has only to be declared as survival time data once.

To get tabular output and a graph of the Nelson-Aalen estimate use the commands

To include 95% confidence limits on the plot use the commands.

To get tabular output and a graph of the Nelson-Aalen estimate use the commands.

Each data set has only to be declared as survival time data once.

Sketch of arguments used to derive Kaplan-Meier's estimate and Nelson-Aalen's estimate

Divide the follow-up period up into a number of small intervals of length \( D_t \) such that each interval contains at most one observation.

The probability \( q_j \) is estimated by

\[
q_j = \frac{1}{b} \sum_{i=1}^{b} \left( \frac{j(i)}{s(i)} \right) \int_0^s d(s) \gamma_j
\]

and

\[
s_j = \frac{1}{b} \prod_{i=1}^{b} \left( \frac{j(i)}{s(i)} \right)
\]

Since

\[
\left( \frac{j(i)}{s(i)} \right)
\]

otherwise

\[
= \frac{1}{b}
\]

uncensored observation

The probability \( q_j \) is estimated by

Divide the follow-up period up into a number of small intervals of length \( D_t \) such that each interval contains at most one observation.

To include 95% confidence limits on the plot use the commands.

Each data set has only to be declared as survival time data once.
KAPLAN-MEIER'S AND NELSON-AALEN'S
ESTIMATES: UNCERTAINTY OF THE ESTIMATES

Estimated standard errors when the data have
no ties among uncensored observations:

\[ \hat{SE}(\hat{S}(t)) = \frac{(\hat{S}(t) - 1)}{\hat{S}(t)} \]

and

\[ q = \frac{1}{\hat{S}(t)} \]

\[ \hat{SE}(\hat{S}(t)) = \sqrt{\hat{S}(t) \cdot SE(\hat{S}(t))} \]

where

\[ \hat{S}(t) \leq \hat{S}(t) \leq \hat{S}(t+1) \leq \hat{S}(t+1) \leq \hat{S}(t+1) \]

A confidence interval of the form:

Improved version (Asymmetric):

close to 0 or close to 1.

Symmetric confidence interval not reasonable for \( \hat{S}(t) \).

(\( \hat{S}(t) - 1 \)) \cdot SE(\( \hat{S}(t) \)) \leq (\( \hat{S}(t) \)) \leq (\( \hat{S}(t) \)) \leq (\( \hat{S}(t) \)) \leq (\( \hat{S}(t) \))

Standard version (Symmetric):

Probability \( \hat{S}(t) \) for fixed \( t \)

Approximate confidence intervals for the survival

CONFIDENCE INTERVALS FOR \( S(t) \)

Approximate confidence intervals for \( S(t) \):

Confidence intervals for both survival times and

Estimates: Uncertainty of the Estimates

Estimated standard errors when the data have no ties.

Note:

Unlike the other estimates of the standard error the
simple estimate for \( S(t) \) changes

both at survival times and censored times. This quantity
has no range restrictions.

Note:

STATA uses a slightly different version:

\[ \hat{SE}(\hat{S}(t)) = \frac{1}{\hat{S}(t)} \]

Simple estimate requires only calculation of

\[ \hat{S}(t) \]

Note:

Unlike the other estimates of the standard error the
simple estimate for \( S(t) \) changes

both at survival times and censored times. This quantity
has no range restrictions.
Confidence Intervals for \( S(t) \): Illustration

The example

\[ \hat{S}(81) = 0.733 \]

\[ \hat{S}(81) = 0.125 \]

\[ \hat{S}(81) = 0.294 \]

\[ \hat{S}(81) = 0.171 \]

Symmetric interval:

Lower bound = \( \hat{S}(81) - 1.96 \times \text{SE}(\hat{S}(81)) \)

Upper bound = \( \hat{S}(81) + 1.96 \times \text{SE}(\hat{S}(81)) \)

Symmetric interval:

Lower bound = \( \hat{S}(81) - 1.96 \times 0.125 \)

Upper bound = \( \hat{S}(81) + 1.96 \times 0.125 \)

Alternative version (asymmetric):

Lower bound = \( \exp(-1.96 \times 0.125) \)

Upper bound = \( \exp(1.96 \times 0.125) \)

Estimate of median survival time \( M \)

Choose \( M \) such that \( \hat{S}(M) = 0.5 \)

An approximate 95\% confidence interval for the median can be obtained from the confidence intervals for the survival probability:

Calculation of 95\% confidence interval for the median (or other percentiles) with STATA: use the command: 95\% CI for survival fct.
METHODS FOR GROUPED FOLLOW-UP DATA

Sometimes exact follow-up times are not available: if e.g. the status of the individuals are ascertained only periodically.

With large data sets it may also be convenient to divide the time scale into a number of categories and use only information about the categories, and not the exact times, in the analysis.

**Estimation of the survival function for grouped data:**

The life table method

**Example:**

Follow-up times in days:

<table>
<thead>
<tr>
<th>55</th>
<th>61+</th>
<th>74</th>
<th>81</th>
<th>93+</th>
<th>122+</th>
<th>138</th>
<th>151</th>
<th>168</th>
<th>202+</th>
<th>220+</th>
<th>238</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>60</td>
<td>120</td>
<td>180</td>
<td>240</td>
<td>300+</td>
<td>360+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consider 60 days intervals: [0,60[, [60,120[, [120,180[, [180,240[.

Let $\ell_i = (\ell_{i-1}, \ell_i)$ denote the cutpoints of the time intervals.

### Estimation

1. **The number alive and at risk at the start of the interval:** $Y = Y^0$.
2. **The number dying in the interval:** $d = \frac{Y \cdot \ell_i}{\ell_{i-1}}$.
3. **The number censored in the interval:** $c = \frac{Y^0}{\ell_{i-1}}$.
4. **The modified number at risk:** $Y^c = Y - c/2$.
5. The conditional death proportion: $q = \frac{d}{Y^c}$.
6. The conditional survival proportion: $p = 1 - q$.

Then compute

\[ S = \frac{Y^0}{Y^c} \]

and finally the estimated survival function

\[ \hat{S}(t) = \prod_{i=1}^{k} \left( 1 - \frac{d_i}{Y^c_i} \right) \]

etc.

\[ S(\ell_j) = S(\ell_{j-1}) \cdot \frac{\ell_j - \ell_{j-1}}{\ell_j} \]

\[ S(\ell_{j-1}) = S(\ell_j) \cdot \frac{\ell_j - \ell_{j-1}}{\ell_j} \]

\[ S(\ell_{j-1}) = S(\ell_j) = \frac{Y^0}{Y^c} \]


**Standard error of the life table estimate:**

Greenwood's formula:

\[
\text{SE} (S(t)) = \sqrt{\frac{\hat{p}(t)}{S(t)} \hat{p}(t) + \hat{p}'(t)}
\]

Example:

At 60 days:
\[
\text{SE} (S(60)) = \sqrt{0.083 \times 0.083 + 0.200} = 0.12
\]

At 120 days:
\[
\text{SE} (S(120)) = \sqrt{0.083 \times 0.083 + 0.200} = 0.32
\]

etc.

**Using Stata for the analysis**

To get the life table estimate of the survival function using Stata, submit one of the commands:

```
Mouse:
Statistics - Survival analysis – Summary statistics, tests & tables – Life tables for survival data
```

To get the life-table estimate of the survival function, use Greenwood's formula:

\[
\text{SE} (S(t)) = \sqrt{\frac{\hat{p}(t)}{S(t)} \hat{p}(t) + \hat{p}'(t)}
\]

Example:

At 60 days:
\[
\text{SE} (S(60)) = \sqrt{0.083 \times 0.083 + 0.200} = 0.12
\]

At 120 days:
\[
\text{SE} (S(120)) = \sqrt{0.083 \times 0.083 + 0.200} = 0.32
\]

etc.
COMPARING SURVIVAL IN TWO GROUPS

Standard non-parametric tests cannot be applied when the data contain censored observations. Several non-parametric procedures, e.g. Wilcoxon-Mann-Whitney rank test, Kruskal-Wallis test, have been generalized to allow for censored observations. Here: primarily log rank test for comparison of survival in two (or several) groups.

The terminology is not very consistent:

- Log rank test
- Generalized Savage’s test
- Mantel-Cox test
- Mantel-Haenszel test for survival data.

Note: Several slightly different versions of these procedures exist. Differences are typically the result of applying different variance estimates.

We shall only consider unpaired two-sample problems here.

Note: The methods to be described are suitable for detection of consistent deviations from the null hypothesis, i.e. if different variance estimates are applied.

Null hypothesis: Same survival in the two groups.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>$s_i$</td>
</tr>
<tr>
<td>$N$</td>
<td>$S$</td>
</tr>
</tbody>
</table>

The terminology is not very consistent:

- Mantel-Haenszel test
- Generalized Savage’s test
- Log rank test
- Cox test

Data: Two samples of survival times, some of which may be censored. Comparison of the survival in two groups of patients.

Problem: Comparing survival in two groups with censored data.

Note: Several non-parametric procedures, e.g. Wilcoxon-Mann-Whitney rank test, are suitable for survival observations that have been observed in two (or several) groups. Primarily log rank test for comparison of survival in two or several groups. The data contain censored observations. The standard non-parametric test cannot be applied when
The test statistics are particularly well suited (i.e. have high power), if the hazard rates are proportional:

\[ \text{hazard rate ratio} = q \]

The parameter \( q \) is called the hazard rate ratio or the mortality rate ratio.

Proportional hazard rates corresponds to the following relations between the survival functions:

\[ S(t)/S(0) = q^t \]

In both plots, \( y(t) \) can be replaced by \( \log y(t) \).

This plot will show two curves with roughly constant vertical distance if the mortality rates are proportional.

The log rank test leads to plots of \( \log \left( \frac{S(t)}{S(0)} \right) - t^*_j \) against \( t \) in the same plot.

It is advisable always to study diagnostic plots, e.g.

1. \( y(t)^*_j \) versus \( y(t^*_j)^*_j \)
2. \( \log(\log y(t)) \) versus \( \log y(t) \)

The points in plot will approximate a straight line if the hazard rates are proportional.

Data: Relapse-free survival in weeks.

### Example 1
A randomized, placebo-controlled trial to assess the effect of remission maintenance therapy with 6-MP on the duration of steroid-induced remission in acute leukemia.

The study was designed and originally analyzed as a matched pairs sequential trial. For illustrative purposes, the data are here analyzed as a standard parallel group matched pairs sequential trial.

<table>
<thead>
<tr>
<th>6-MP Group</th>
<th>Placebo Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 11+ 25+ 1 5 12</td>
<td>6 13 32+ 1 8 15</td>
</tr>
<tr>
<td>6 16 32+ 2 8 17</td>
<td>6+ 17+ 34+ 2 8 22</td>
</tr>
<tr>
<td>7 19+ 35+ 3 8 23</td>
<td>9+ 20+ 4 11</td>
</tr>
<tr>
<td>10 22 4 11</td>
<td>10+ 23 5 12</td>
</tr>
</tbody>
</table>

The test statistics are not well suited in situations, where the hazard rates cross (this will for instance be the case if the survival functions cross):

The points in plot will show two curves with roughly constant vertical distance if the mortality rates are proportional.
The relapse rates seem roughly proportional.

Estimated survival curves:

- Placebo
- 6-MP

Time in weeks

Estimated integrated hazards:

- Placebo
- 6-MP

Time in weeks

Diagnostic plots:

- Placebo Integrated hazard
- 6-MP Integrated hazard

Log integrated hazard

Time in weeks

Integrated hazard

Estimated survival curves:

- Placebo
- 6-MP

Time in weeks

Probability

The relapse rates seem roughly proportional.
A valid test can be established by comparing the pattern of occurrences of relapses in the two groups controlling for the time at which the relapses occur. This approach resembles the Mantel-Haenszel test, but the formal justification is different.

**Basic Idea:**

For each event-time a $2 \times 2$ table is established:

<table>
<thead>
<tr>
<th>Event time</th>
<th>6-MP</th>
<th>Placebo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

At risk at time $t$ the proportion of patients at risk belonging to group 1 is $\frac{Y_{1t}}{Y_{t}}$.

If the relapse rates are identical we will therefore expect the same proportion of the relapses to occur in group 1, i.e. the expected number in group 1 is $\frac{Y_{1t}}{Y_{t}} \times \text{Total}$.

A total of 9 relapses is observed in the treatment group. On the null hypothesis we would have expected 19.25 relapses based on the relative size of the two groups at the 17 different relapse times.

A variance estimate is needed to assess the significance of this discrepancy.
A valid variance estimate can be obtained as the sum of contributions from each \( 2 \times 2 \) table in the same way as for the Mantel-Haenszel procedure. For the first table

<table>
<thead>
<tr>
<th>Event</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
<td>19</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>21</td>
<td>42</td>
</tr>
</tbody>
</table>

The contribution to the variance becomes

\[
V = \sum \left( \frac{Y(i) - m(i)}{w(i)} \right)^2
\]

For the first table

\[
\begin{align*}
V & = \sum \left( \frac{Y(i) - m(i)}{w(i)} \right)^2 \\
& = \left( \frac{0 - 1}{2} \right)^2 + \left( \frac{2 - 2}{2} \right)^2 \\
& = \frac{1}{4} + 0 \\
& = 0.25
\end{align*}
\]

Calculating the remaining 16 contributions in a similar fashion and adding the results together gives

\[
V = \sum V = 6.257
\]

The test statistic, the log rank test, is obtained as

\[
lr = \frac{\sum (\text{Obs.} - \text{Exp})^2}{Variance} = \frac{6.257}{9.251} = 0.679
\]

On the null hypothesis of identical survival distributions, the distribution of the test statistic is approximately a \( \chi^2 \) distribution with 1 degree of freedom. Large values provide evidence against the null hypothesis. A value of 16.79 is highly significant (\( p = 0.0004 \)).

The log rank test – in general

Consider two samples of possibly censored survival times. The sample sizes are denoted as \( n_1 \) and \( n_2 \). Let \( d \) denote the number of distinct values found among the uncensored observations in the combined sample and among the censored and uncensored values that are ordered by taking uncensored values first, then among censored and uncensored values, and ending order. The uncensored values are sorted in ascending order. The log rank test statistic is obtained as

\[
lr = \sum \frac{d(i)Y(t)1 - m(t)}{w(t)}
\]

For the first table

<table>
<thead>
<tr>
<th>Event time ( t )</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
<td>19</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>21</td>
<td>42</td>
</tr>
</tbody>
</table>

The log rank test statistic is

\[
lr = \sum \frac{d(i)Y(t)1 - m(t)}{w(t)}
\]

For the Mantel-Haenszel procedure, contributions from each \( 2 \times 2 \) table in the same way as for the
The expected number and variance are computed on the hypothesis of identical hazard rates. The contributions from each table are added together:

\[ \sum_{i=1}^{d} \left( O_i - E_i \right)^2 \]

where the observed and expected numbers are those described above. The alternative version of the log rank test is slightly conservative, since the variance increases with the number of ties among event times the discrepancy is minimal but the test is therefore too large if the data contain no tied events.

The log rank test is finally obtained as

\[ X^2 = \sum_{i=1}^{d} \frac{O_i - E_i}{E_i} \]

On the null hypothesis \( X^2 \) is approximately a \( \chi^2 \) variate on 1 degree of freedom. Large values provide evidence against the null hypothesis.

Comments:

1. Exactly the same test statistic is obtained if we instead consider group 2 and compute the observed and expected number of deaths in this group.

2. Some computer packages and text books use the name log rank test for a slightly different test statistic

\[ \sum_{i=1}^{d} \left( O_i - E_i \right)^2 + \sum_{i=1}^{d} \left( E_i - E_i \right)^2 = X^2 \]

On the null hypothesis \( X^2 \) is approximately a \( \chi^2 \) variate on 1 degree of freedom. Large values provide evidence against the null hypothesis.

3. The alternative version of the log rank test is (slightly) conservative. The standard terminology "observed" and "expected" values in the usual sense are not the expected values of conditional expectations. They represent our expectation based on the relative size of the two groups at the different event times. The standard terminology "observed" and "expected" values in the usual sense is appealing, but not really justified. "Expected" and "expected" values are not equivalent.

The expected numbers of events are those described above. The alternative version of the log rank test is (slightly) conservative, since the expected values are the sums of conditional expectations. The alternative version of the log rank test is (slightly) conservative, since the expected values are the sums of conditional expectations. The alternative version of the log rank test is (slightly) conservative, since the expected values are the sums of conditional expectations.
USING STATA FOR THE ANALYSIS

Read data into Stata and declare data as survival time data. The data set has three variables: time, failure, and treatm.

To compare the survival in the two treatment groups using log rank test write

```
sts test treatm, noshow
```

The first two lines are suppressed if the option noshow is added.

```
<table>
<thead>
<tr>
<th>time</th>
<th>failure</th>
<th>treatm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>30.00</td>
<td></td>
</tr>
<tr>
<td>1.12</td>
<td>22</td>
<td></td>
</tr>
</tbody>
</table>
```

Log-rank test for equality of survival functions

```
arrive = d: status == 1
```

```
OUTPUT
```

```
sts test treatm
```

```
sts test treatm using log rank test with
```

To compare the survival in the two treatment groups

```
use E:\kurser\survival\carthage2003\ex33.dat
```

```
frame data. The data set has three time variables: time, failure, and
```

Read data info Stata and declare data as survival time

```
USING STATA FOR THE ANALYSIS
```

```
TABLE output:
```

```
sts graph, na cna by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```

```
sts graph, na by(treatm)
```
ALTERNATIVE NON-PARAMETRIC TEST

The log rank test is essentially a scaled version of the difference $\frac{O_i - E_i}{\sqrt{O_i + E_i}}$. Since

The contribution to the numerator from each $2 \times 2$ has the form $(O_i - E_i) \cdot \frac{\hat{r}}{\hat{r}} = (O_i - E_i) \cdot \frac{\hat{r}}{\hat{r}}$.

The log rank test is essentially a scaled version of the

STATA output

The option `cox` can be used in a proportional hazards model to compute with additional option: The likelihood ratio test (more on that)

Wilcoxon test is the data contains no censored observations.

Wilcoxon test if the data contains no censored observations.

STATA: 

1. sts test, p moscow

The test statistic has approximately a $\chi^2$ distribution on 1 degree of freedom on the null hypothesis.

In STATA the following test statistics are available.

By using different weighting schemes alternative test statistics are obtained. All these tests have the form given equal weight.

Additional option: The likelihood ratio test is computed with the option `cox`.

Additional option: The conditional test is performed in the same way as the log rank test (more on that).
DESCRIBING EXCESS MORTALITY

The log rank test and the alternative test statistics are non-parametric tests used to assess if chance fluctuations are a likely explanation for an observed discrepancy between two survival curves. Generally, it is also important to describe such a difference and quantify it in a suitable way.

Plotting the two Kaplan-Meier estimates to be compared in the same figure is one possibility. But it is often convenient to supplement such a figure with a simple numerical description of the difference in mortality.

Possible summaries include:

1. The difference or the ratio of e.g. the two-year survival probabilities.
2. The difference or the ratio of the estimated median survival times.
3. An estimate of the ratio of the mortality rates. If the mortality rates are approximately proportional this corresponds to estimating the constant of proportionality \( q \) (see page 31).

The mortality rate ratio as a measure of discrepancy

An estimated mortality rate ratio (summary measure 3) could be presented e.g. as

\[
\text{"Treatment A reduces the mortality with 30\% relative to treatment B"},
\]

i.e.,

\[
\frac{\hat{q}}{\hat{q}} = 0.70, \text{ often described as the relative risk.}
\]

The regression coefficients computed by a Cox's regression analysis (to be discussed later) can after a simple transformation (taking exp) be interpreted as estimated mortality rate ratios.

Occasionally, it is necessary to use alternative estimates of the mortality rate ratio. This could for instance be the case if access to the original data is not possible, e.g. in a meta-analysis. Several proposals exist, e.g.:

A "quick and dirty" estimate of \( q \) could be computed as

\[
\hat{q} = \frac{O}{E} \times \frac{E}{O} = \hat{q}
\]

where \( O \) and \( E \) are the observed and expected numbers of events, respectively. This estimate is known to be bias towards 1, but the regression coefficient of the log hazard ratio in the Cox's regression model can be interpreted as the estimated mortality rate ratio.

Generally, it is also important to describe a difference and quantify it in a suitable way. The log rank test is one possibility, but it is often non-parametric tests used to assess if chance fluctuations are a likely explanation for an observed discrepancy between two survival curves. The alternative test statistics are.
Example (Freireich et al's data)
The output from the command sts test allow calculation of the "quick and dirty" estimate.

```
    test 0.001 1.5 2.1 2.6338 4.52 0.5823
```

The output shows that the optimal estimate is 4.52 with a confidence interval of 2.6338 to 4.52.

Confidence intervals are not computed by sts test (use stcox instead).

THE K-SAMPLE PROBLEM: COMPARING SEVERAL SURVIVAL CURVES

Problem: Comparison of the survival in \( K \) (\( K \geq 2 \)) different populations.

Data: A sample of possibly censored survival times from each population.

Null hypothesis: \( \lambda_i = \lambda_j \) for all \( i, j \in \{1, \ldots, K\} \).

Test statistics: The test statistics used for comparing two groups can all be generalized to comparison of more than two groups.

STATA: \( K \)-sample log rank test of age effects on mortality in patients with malignant melanoma.

```
sts test agep < agep
```

The output from the command sts test allow calculation of the "quick and dirty" estimate.

```
    test 0.001 1.5 2.1 2.6338 4.52
```

The output shows that the optimal estimate is 4.52 with a confidence interval of 2.6338 to 4.52.

Confidence intervals are not computed by sts test (use stcox instead).
THE K SAMPLE PROBLEM WITH ORDERED CATEGORIES: TEST FOR TREND.

Problem: Again, comparison of the survival in different groups of patients, but there is a "natural" ordering of the groups, they might e.g. represent different age groups, different dose levels of some exposure or increasing severity of the disease at time of entry.

Data: A sample of possibly censored survival times from each population.

Null hypothesis: The exposure of increasing severity of the disease at time of entry.

For all $i$: $\gamma_i = \cdots = \gamma_i \neq (i)$

Test: log rank test.

The log rank test is a normalized version of the statistic

$$\left(\sum_{j=1}^{K} \frac{X_{ij}}{\sum_{j=1}^{K} X_{ij}} \cdot \left( O_j - E_j \right)^2 \right) \cdot (\sum_{j=1}^{K} \frac{1}{\sum_{j=1}^{K} X_{ij}})^{-1}$$

where $O_j$ and $E_j$ are the observed and expected numbers derived as described for the $K$ sample log rank test and $X_j$ are the group scores, i.e. values assigned to the groups reflecting the ordering (e.g. average age of patients in the group or dose level).

$O_j$ and $E_j$ are 'observed and expected numbers derived as described for the $K$ sample log rank test and $X_j$ are the group scores, i.e. values assigned to the groups reflecting the ordering (e.g. average age of patients in the group or dose level).

$T$ is an estimate of the variance of the numerator. On the null hypothesis the trend test is approximately distributed as a $\chi^2$ variate on 1 degree of freedom. Large values provide evidence against the null hypothesis. Large values provide evidence against the null hypothesis.

Solution: For each of the $K$-sample tests a corresponding test for trend has been developed.

$T$ is an estimate of the variance of the numerator. On the null hypothesis the trend test is approximately distributed as a $\chi^2$ variate on 1 degree of freedom. Large values provide evidence against the null hypothesis.

The log rank test for trend:

$$\frac{\sum_{j=1}^{K} \frac{X_{ij}}{\sum_{j=1}^{K} X_{ij}} \cdot \left( O_j - E_j \right)^2 \cdot (\sum_{j=1}^{K} \frac{1}{\sum_{j=1}^{K} X_{ij}})^{-1}}{\sum_{j=1}^{K} \frac{1}{\sum_{j=1}^{K} X_{ij}}}$$

The usual $K$-sample tests (log rank test etc.) are valid tests, but not particularly well suited for detecting this type of departures from the null hypothesis. Information about the ordering is not utilized: The value of these test statistics does not change if the order of the groups is changed.

Solutions: For each of the $K$-sample tests a corresponding test for trend has been developed.

$T$ is an estimate of the variance of the numerator. On the null hypothesis the trend test is approximately distributed as a $\chi^2$ variate on 1 degree of freedom. Large values provide evidence against the null hypothesis.
is derived from these accumulated statistics. Note:

A linear transformation of the scores will not change the value of the test statistic, e.g.

<table>
<thead>
<tr>
<th>Group</th>
<th>Scores</th>
<th>Alt. scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>145</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>190</td>
<td>1</td>
</tr>
</tbody>
</table>

The two sets of scores give the same value of the test statistic, but the alternative scores are much simpler to use.

**Example:** log rank test for trend with age in mortality of patients with malignant melanoma.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Observed</th>
<th>Expected</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>145</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>190</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The two sets of scores give the same value of the test statistic, e.g.

A linear transformation of the scores will not change the value of the test statistic, e.g.
A STRATIFIED ANALYSIS BASED ON THE TWO-SAMPLE LOG RANK TEST:

**Problem:**
Comparison of the survival after two different treatment modalities A and B controlling for the effect of the age of the patients, say.

**Data:**
Two samples of lifetimes, some of which may be censored. The value of the factor, which is used to define the strata, must also be known for each patient.

**Procedure:**
1. The age range of patients are divided up into a suitable number of age categories. Patients belonging to each age group are considered separately.
2. For each stratum the survival in the two treatment groups is summarized by the observed number of deaths and the corresponding expected numbers and the variance. These quantities are all derived in the same way as for the usual log rank test.
3. The observed and expected numbers and the variance are summed across strata and a test statistic is computed from these sums.

Stratum 1:
\( \text{Age} < 30 \).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Observed</th>
<th>Expected</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stratum 2:
\( 30 \leq \text{Age} < 50 \).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Observed</th>
<th>Expected</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stratum 3:
\( 50 \leq \text{Age} < 70 \).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Observed</th>
<th>Expected</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stratum 4:
\( \text{Age} \geq 70 \).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Observed</th>
<th>Expected</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summed across strata:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Observed</th>
<th>Expected</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>( \sum )</td>
<td>( \sum )</td>
<td>( \sum )</td>
</tr>
<tr>
<td>B</td>
<td>( \sum )</td>
<td>( \sum )</td>
<td>( \sum )</td>
</tr>
</tbody>
</table>

The stratified log rank test is based on these statistics: variances are summed across strata and a test statistic is computed from these sums.
1. The stratification may be based on several variables. Increasing the number of variables to use will typically improve the comparability within strata (tighter control), but at the same time it may reduce the effective sample size: The risk of having strata with observations from only one of the groups increases also with the number of strata, and such strata can not be utilized in the analysis.

2. Because too many strata may reduce the effective sample size, one will typically not include among the stratifying factors risk factors with approximately the same distribution in the groups.

3. Note that the amount of computation depends largely on the number of events, so a stratified analysis requires roughly the same computational effort as the corresponding unstratified test.

4. A Cox regression analysis will allow to control for effects of confounding factors when assessing the difference between the survival in two treatment groups.

5. A Cox regression analysis will also allow a study of the treatment difference between strata (interaction). The stratified analysis implicitly assumes that a treatment difference has variation in the treatment difference between strata.

STRATIFIED ANALYSIS USING STATA

A log rank test to compare the survival in two exposure groups (`expa`) stratified by `agecat` and `sex`:

```
sts test expa, strata(agecat sex)
```

A Peto-Prentice test for trend in age (`agecat`) stratified by `sex`:

```
sts test expa, strata(agecat sex) detrend
```

The stratification is accomplished by `agecat` which is a variable that has been created from the original `age`. The stratification variable can be any variable that gives meaningful strata for the analysis.

```
by sex with detrended output
```

A log rank test for trend in age (agecat) stratified by sex:

```
sts test expa, strata(agecat group)
```

```
egen agecat=group(agecat sex) detrend
```

Stratified versions of the other non-parametric tests are available as well. The basic idea is in all cases to compute the summary statistics within each stratum, add them together across strata, and compute the relevant test statistic. For example, the basic idea is in all cases to compute the Wilcoxon statistic, and then for each stratum, compute the Wilcoxon statistic under the null hypothesis for that stratum.

```
sts test expa, strata(agecat group)
```

The stratified version of the other non-parametric tests are:

1. The stratification may be based on several variables.
2. Because too many strata may reduce the effective sample size, one will typically not include among the stratifying factors risk factors with approximately the same distribution in the groups.
3. Note that the amount of computation depends largely on the number of events, so a stratified analysis requires roughly the same computational effort as the corresponding unstratified test.
4. A Cox regression analysis will allow to control for effects of confounding factors when assessing the difference between the survival in two treatment groups.
5. A Cox regression analysis will also allow a study of the treatment difference between strata (interaction). The stratified analysis implicitly assumes that a treatment difference has variation in the treatment difference between strata.

STRATIFIED ANALYSIS BASED ON THE OTHER NON-PARAMETRIC TESTS:

Stratified versions of the other non-parametric test statistics and of test for trend can be computed in a similar fashion. The basic idea is in all cases to compute the test statistic for each stratum, add them together across strata, and compute the relevant test statistic from these sums.

STRATIFIED ANALYSIS USING STATA

A log rank test to compare the survival in two exposure groups (`expa`) stratified by `agecat` and `sex`:

```
sts test expa, strata(agecat sex)
```

A Peto-Prentice test for trend in age (`agecat`) stratified by `sex`:

```
sts test expa, strata(agecat sex) detrend
```

The stratification is accomplished by `agecat` which is a variable that has been created from the original `age`. The stratification variable can be any variable that gives meaningful strata for the analysis.

```
by sex with detrended output
```

A log rank test for trend in age (agecat) stratified by sex:

```
sts test expa, strata(agecat group)
```

The stratified version of the other non-parametric tests are available as well. The basic idea is in all cases to compute the summary statistics within each stratum, add them together across strata, and compute the relevant test statistic. For example, the basic idea is in all cases to compute the Wilcoxon statistic, and then for each stratum, compute the Wilcoxon statistic under the null hypothesis for that stratum.

```
sts test expa, strata(agecat group)
```

The stratified version of the other non-parametric tests are:

1. The stratification may be based on several variables.
2. Because too many strata may reduce the effective sample size, one will typically not include among the stratifying factors risk factors with approximately the same distribution in the groups.
3. Note that the amount of computation depends largely on the number of events, so a stratified analysis requires roughly the same computational effort as the corresponding unstratified test.
4. A Cox regression analysis will allow to control for effects of confounding factors when assessing the difference between the survival in two treatment groups.
5. A Cox regression analysis will also allow a study of the treatment difference between strata (interaction). The stratified analysis implicitly assumes that a treatment difference has variation in the treatment difference between strata.