STATISTICAL ANALYSIS OF SURVIVAL DATA IN CLINICAL RESEARCH 1

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 event in time. Events are experienced by (independent) individuals.

Standard setting: At most one event per individual.

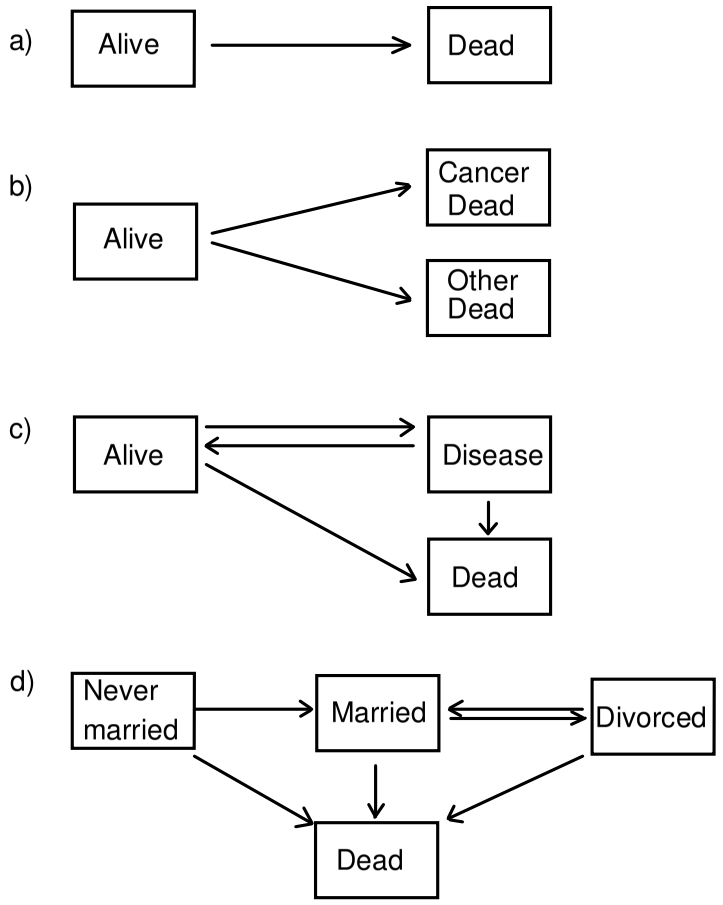
Some examples:
- Chronic disease incidence
- Mortality

Also:
- Product reliability
- Life insurance
- Duration of unemployment

More generally: Several types of events
- *Example:* Cause-specific mortality

More generally: Several events per individual
- *Example:* Recurrent diseases

Some standard situations



SURVIVAL DATA

Two aspects: ***A waiting time***
Occurrence of an event

A continuous and a discrete part.

Choice of time scale:
A waiting time, but what is the appropriate starting point?

- *Time since entry*
- *Birth, i.e. time = age*
- *Time since diagnosis*

or something else?

Describing occurrence in time of a health-related event:

How do we **quantify**

Prognosis,
Risk,
Chance ?

Probabilities are an obvious choice, but not the only possibility.

Compare: **Is it dangerous to travel by airplane ?**

DATA: For each individual in the study:

A waiting time (event occurred) or an ***incomplete*** waiting time (no event occurred).

i.e. each individual contributes:

(t,d) = (***period of observation***, ***status*** at the end of the period)

Status:

$d=1$	The period of observation is terminated by the occurrence of the event
$d=0$	The event has not (yet) occurred at the end of the period of observation.

(Right) censoring:
Incomplete information: For some patients the exact value of the waiting time is not observed, but it is known that the waiting time exceeds a specified value.

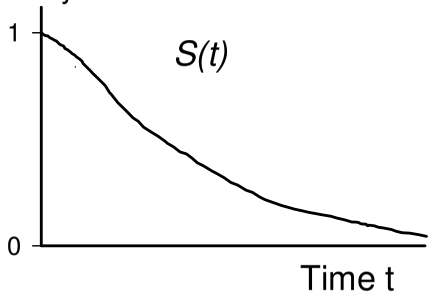
Additional feature, which complicates the data analysis:
In most studies the ***potential follow-up time*** (= time from entry date until closing date) will vary among patients included in the study.

Consequences of ***right censoring*** and ***varying follow-up times:*** **Standard statistical methodology can not be applied directly. Modifications, or generalizations are necessary.**

DESCRIBING THE PROGNOSIS, HOW ?

Survival function = Survivorship function
= Survival curve

$S(t)$ = The probability of being alive (event-free) at time after entry



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***

Notation: several versions in common use

$$m(t) = \lambda(t) = \mu(t) = h(t)$$

The mortality rate describes the instantaneous risk of dying.

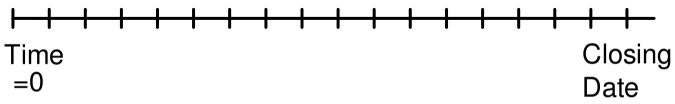
More formally

$m(t) \cdot \Delta t =$ The probability of dying before time $t + \Delta t$ if alive at time t (Δt small)

The mortality rate gives the relative rate of change of the survival function

Estimating the mortality rate:

Divide the follow-up period up in small intervals:



For each small time interval compute:

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

The ratio

$$\frac{\text{Death}}{\text{Risk time}} = \frac{\text{Number of deaths in the interval}}{\text{Total time lived in the interval}}$$

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

Note:

$S(t)$	A probability, i.e. No dimension	$0 \le S(t) \le 1$
$m(t) = \lambda(t)$	A rate, i.e. Dimension = per time unit	$0 \le m(t)$

Relations between the theoretical functions

$S(t) =$ Survival function

$\lambda(t) =$ Mortality rate, Death intensity

$\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s =$ Integrated mortality rate or integrated hazard

$$S(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(s)\,\mathrm{d}s\right)$$

$$\Lambda(t) = -\ln(S(t))$$

$$\lambda(t) = \frac{d\,\Lambda(t)}{dt} = -\frac{d}{dt}\ln(S(t)) = -\frac{S'(t)}{S(t)} = -\frac{\frac{d}{dt}S(t)}{S(t)}$$

i.e.

$$\lambda(t) = \frac{\left(S(t) - S(t+dt)\right)/dt}{S(t)}$$

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

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 ?

If yes When ?
If no Event-free for how long ?

Basic data summaries:

$Y(t)$ = Number of patients alive and not censored just before time t = **Number at risk at time *t***.

$N(t)$ = Number of patients with survival times less than or equal to t = Number of deaths up to and including time t.

$Y(0) = n$ and $N(0) = 0$

If data contain no censored observations then

$Y(t)$ = $n \cdot$ empirical survival function

$N(t)$ = $n \cdot$ empirical distribution function

With censored data these relations are no longer true.

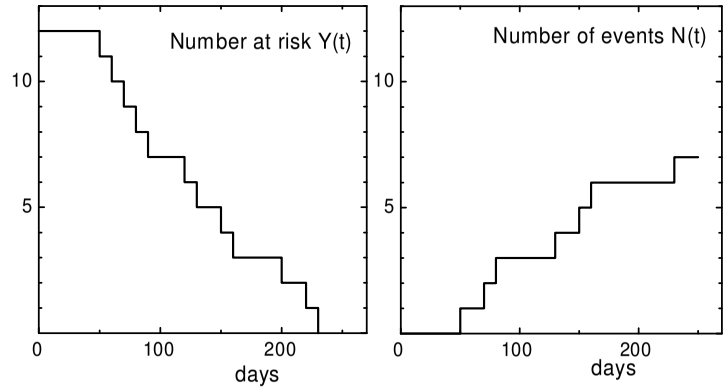
A simple example

Data:

Survival time in days measured from time of entry
(Censored observations marked by +)

55	61+	74	81	93+	122+
138	151	168	202+	220+	238

Sample size = n = 12



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.

In the example

Time	At Risk
55	12
61+	11
74	10
81	9
93+	8
122+	7
138	6
151	5
168	4
202+	3
220+	2
238	1

2) For each time of death t_i compute $Y_i = Y(t_i)$ and then

$$q_i = 1/Y_i \qquad p_i = 1 - q_i = 1 - 1/Y_i$$

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

$\hat{S}(t)$ =the products of the p_i's from death times $\leq t$

The example:

$$\hat{S}(55) = \frac{11}{12} = 0.917$$

$$\hat{S}(74) = \frac{11}{12}\cdot\frac{9}{10} = 0.825$$

$$\hat{S}(81) = \frac{11}{12}\cdot\frac{9}{10}\cdot\frac{8}{9} = 0.733$$

$$\hat{S}(138) = \frac{11}{12}\cdot\frac{9}{10}\cdot\frac{8}{9}\cdot\frac{5}{6} = 0.611$$

$$\hat{S}(151) = \frac{11}{12}\cdot\frac{9}{10}\cdot\frac{8}{9}\cdot\frac{5}{6}\cdot\frac{4}{5} = 0.489$$

$$\hat{S}(168) = \frac{11}{12}\cdot\frac{9}{10}\cdot\frac{8}{9}\cdot\frac{5}{6}\cdot\frac{4}{5}\cdot\frac{3}{4} = 0.367$$

$$\hat{S}(238) = \frac{11}{12}\cdot\frac{9}{10}\cdot\frac{8}{9}\cdot\frac{5}{6}\cdot\frac{4}{5}\cdot\frac{3}{4}\cdot\frac{0}{1} = 0.000$$

KAPLAN-MEIER'S ESTIMATE

Data: A sample of size *n.* The number of observed deaths *d (d ≤ n).* In the example $n = 12$ and $d = 7$.

Ordered survival times (uncensored observations).

$$t_1 < t_2 < \cdots < t_d$$

Kaplan-Meier's estimate:

$$\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{1}{Y(t_i)} \right)$$

A product: Each time of death $t_i \le t$ contributes a factor to this product.

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

If ties are present in the data:

$$\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{m_i}{Y(t_i)} \right)$$

where m_i = number of deaths at time t_i

If censored and uncensored observations are tied:
Deaths occur before censoring

No censored observations:
$\hat{S}(t) = 1 -$ empirical distribution function

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. *time* and *status*

time	status
55	1
61	0
74	1
81	1
93	0
122	0
138	1
151	1
168	1
202	0
220	0
238	1

2) Define the data to be survival time data

a) On the command line write

```
stset time , failure(status==1)
```

or use the mouse

b) *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 list
```

or use the mouse: *Statistics - Survival analysis – Summary statistics, tests & tables – List survivor and cumulative hazard function* and fill out the menu.

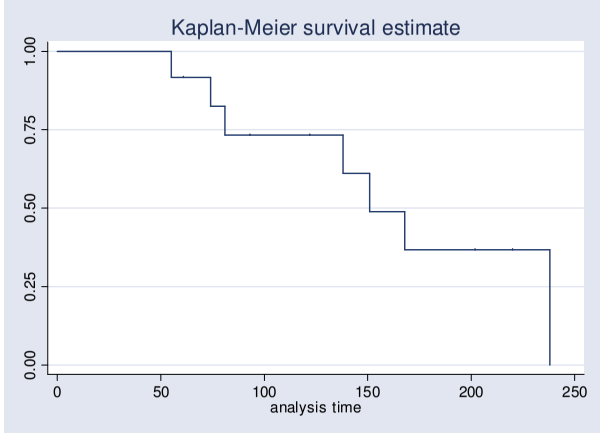
OUTPUT

```
. stset time , failure(status==1)

     failure event:  status == 1
obs. time interval:  (0, time]
 exit on or before:  failure

--------------------------------------------------------------
    12  total obs.
     0  exclusions
--------------------------------------------------------------
    12  obs. remaining, representing
     7  failures in single record/single failure data
  1603  total analysis time at risk, at risk from t =         0
                            earliest observed entry t =         0
                                 last observed exit t =       238

. sts list

         failure _d:  status == 1
   analysis time _t:  time

         Beg.          Net    Survivor    Std.
  Time    Total    Fail Lost   Function   Error     [95% C.I.]
-------------------------------------------------------------------
    55       12       1    0     0.9167  0.0798    0.5390  0.9878
    61       11       0    1     0.9167  0.0798    0.5390  0.9878
    74       10       1    0     0.8250  0.1128    0.4609  0.9533
    81        9       1    0     0.7333  0.1324    0.3790  0.9056
    93        8       0    1     0.7333  0.1324    0.3790  0.9056
   122        7       0    1     0.7333  0.1324    0.3790  0.9056
   138        6       1    0     0.6111  0.1569    0.2546  0.8375
   151        5       1    0     0.4889  0.1664    0.1623  0.7545
   168        4       1    0     0.3667  0.1637    0.0908  0.6574
   202        3       0    1     0.3667  0.1637    0.0908  0.6574
   220        2       0    1     0.3667  0.1637    0.0908  0.6574
   238        1       1    0     0.0000       .         .       .
-------------------------------------------------------------------
```


4) to get a graph as output write on the command line

```
sts graph
```

to get



to include 95% pointwise confidence limi on the graph write

```
sts graph , gwood
```

Some additional options (to be placed after the comma)

```
atrisk
censored
tmin(#)
tmax(#)
```

AN ESTIMATE OF THE INTEGRATED HAZARD: NELSON-AALEN'S METHOD

The calculations involves the following steps:

1) The observations are sorted in ascending order.

2) For each time of death t_i compute $Y_i = Y(t_i)$ and then

$$q_i = 1/Y_i$$

3) Nelson-Aalen's estimate

$\hat{\Lambda}(t)$ =the sum of the q_i's from death times $\leq t$

The example:

$$\hat{\Lambda}(55) = \frac{1}{12} = 0.083$$

$$\hat{\Lambda}(74) = \frac{1}{12}+\frac{1}{10} = 0.183$$

$$\hat{\Lambda}(81) = \frac{1}{12}+\frac{1}{10}+\frac{1}{9} = 0.294$$

$$\hat{\Lambda}(138) = \frac{1}{12}+\frac{1}{10}+\frac{1}{9}+\frac{1}{6} = 0.461$$

$$\hat{\Lambda}(151) = \frac{1}{12}+\frac{1}{10}+\frac{1}{9}+\frac{1}{6}+\frac{1}{5} = 0.661$$

$$\hat{\Lambda}(168) = \frac{1}{12}+\frac{1}{10}+\frac{1}{9}+\frac{1}{6}+\frac{1}{5}+\frac{1}{4} = 0.911$$

$$\hat{\Lambda}(238) = \frac{1}{12}+\frac{1}{10}+\frac{1}{9}+\frac{1}{6}+\frac{1}{5}+\frac{1}{4}+\frac{1}{1} = 1.911$$

NELSON-AALEN'S ESTIMATE

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

Nelson-Aalen's estimate:

$$\hat{\Lambda}(t) = \sum_{t_i \leq t} \frac{1}{Y(t_i)}$$

If ***ties*** are present use

1) $$\hat{\Lambda}(t) = \sum_{t_i \leq t} \frac{m_i}{Y(t_i)}$$

or

2) Separate the survival times slightly and use basic formula.

Example: If for some data $Y(t_i) = 105$ and $m_i = 3$ then

Contribution using 1) $\frac{3}{105}$

Contribution using 2) $\frac{1}{105}+\frac{1}{104}+\frac{1}{103}$

Note: $\Lambda(t) = -\ln(S(t))$,
but $\hat{\Lambda}(t) < -\ln(\hat{S}(t))$

The difference is usually small

$$0 < -\ln(\hat{S}(t)) - \hat{\Lambda}(t) < \frac{1}{Y(t)} - \frac{1}{n}$$

USING STATA FOR THE ANALYSIS

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

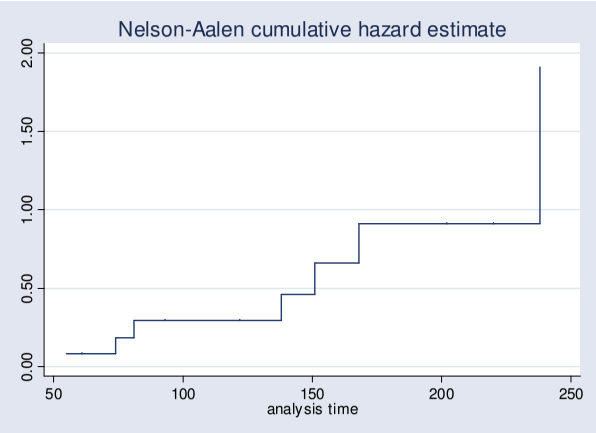
To get tabular output and a graph of the Nelson-Aalen estimate use the commands

```
sts list , na
sts graph , na
```

To include 95% confidence limits on the plot

```
sts graph , na cna
```

OUTPUT (plot only)



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 Δt such that each interval contains at most one observation.

Let t_j denote the lower limit of the j'th interval and let q_j denote the probability of dying in the j'th interval given alive at the start of the interval (at time t_j).

The probability q_j is estimated by

$$\hat{q}_j = \begin{cases} 1/Y(t_j) & \text{if the interval contains an uncensored observation} \\ 0 & \text{otherwise} \end{cases}$$

Since

$$S(t) = \prod_{t_j \le t} (1 - q_j)$$

and

$$\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s \approx \sum_{t_j \le t} \lambda(t_j)\,\Delta t \approx \sum_{t_j \le t} q_j$$

estimates of $S(t)$ and $\Lambda(t)$ can be obtained by replacing the probabilities q_j by their estimated values $\hat{q}_j$.

Note that intervals not containing death times can be omitted. They contribute a factor 1 to the product and a term 0 to the sum.

KAPLAN-MEIER'S AND NELSON-AALEN'S ESTIMATES: UNCERTAINTY OF THE ESTIMATES

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

$$SE(\hat{\Lambda}(t)) = \sqrt{\sum_{t_i \le t} \frac{1}{Y(t_i)^2}}$$

(note: STATA uses a slightly different version) and

$$SE(\hat{S}(t)) = \hat{S}(t) \cdot SE(\hat{\Lambda}(t))$$

If ***ties*** are present

1) $$SE(\hat{\Lambda}(t)) = \sqrt{\sum_{t_i \le t} \frac{m_i}{Y(t_i) \cdot (Y(t_i) - m_i + 1)}}$$

or

2) Separate the survival times slightly and first version.

Simple estimate (requires only calculation of $\hat{S}(t)$):

$$SE(\hat{S}(t)) = \hat{S}(t) \cdot \sqrt{\left(1 - \hat{S}(t)\right) / Y(t)}$$

Note:
Unlike the other estimates of the standard error the simple estimate changes ***both*** at survival times ***and*** censoring times.

CONFIDENCE INTERVALS FOR $S(t)$

Approximate confidence intervals for the survival probability $S(t)$ for fixed t.

Standard version (Symmetric):

$$\hat{S}(t) - 1.96 \cdot SE(\hat{S}(t)) \le S(t) \le \hat{S}(t) + 1.96 \cdot SE(\hat{S}(t))$$

Symmetric confidence interval not reasonable for $S(t)$ close to 0 or close to 1.

Improved version (Asymmetric):

A confidence interval of the form

$$\left(\hat{S}(t)\right)^a \le S(t) \le \left(\hat{S}(t)\right)^b$$

where

$$b = \exp\left(-1.96 \cdot SE(\ln \hat{\Lambda}(t))\right)$$

$$a = 1/b$$

and

$$SE(\ln \hat{\Lambda}(t)) = SE(\hat{\Lambda}(t)) / \hat{\Lambda}(t)$$

The asymmetric confidence interval is obtained from a symmetric confidence interval of $\ln(\Lambda(t))$. This quantity has no range restrictions.

Confidence intervals for S(t): Illustration

The example

$$\hat{S}(81) = 0.733 \qquad SE(\hat{S}(81)) = 0.125$$
$$\hat{\Lambda}(81) = 0.294 \qquad SE(\hat{\Lambda}(81)) = 0.171$$

Symmetric interval:

$$\text{Lower bound} = 0.733 - 1.96 \cdot 0.125 = \quad 0.488$$
$$\text{Upper bound} = 0.733 + 1.96 \cdot 0.125 = \quad 0.978$$

Alternative version (asymmetric):

$$b = \exp\left(-1.96 \cdot 0.171 / 0.294\right) = \exp(-1.140) = 0.320$$

$$a = 1/b = 3.127$$

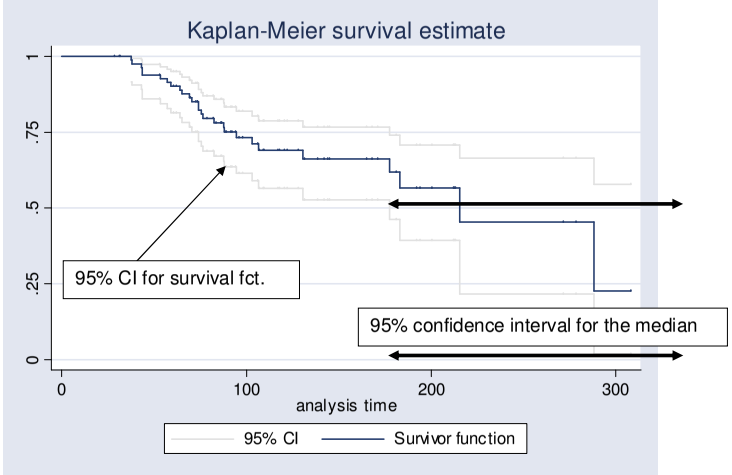
$$\text{Lower bound} = \hat{S}(81)^a = 0.733^{3.13} = 0.379$$
$$\text{Upper bound} = \hat{S}(81)^b = 0.733^{0.32} = 0.905$$

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 `stci`

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.

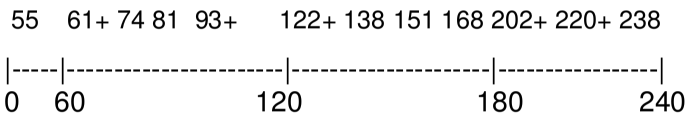
Estimation of the survival function for grouped data: The life table method

Example:

Follow-up times in days:

55	61+	74	81	93+	122+
138	151	168	202+	220+	238

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



Let $\tau_0, \tau_1, \tau_2, \tau_3, \ldots, \tau_k$ denote the cutpoints of the time categories.

For each interval determine

1. $Y_i = Y(\tau_{i-1}) =$ The number alive and at risk at the start of *i'* th interval.

2. $d_i = N(\tau_i -) - N(\tau_{i-1} -) =$ The number dying in the interval.

3. $c_i =$ The number censored in the interval.

Then compute

4. The modified number at risk $n_i = Y_i - c_i/2$

5. The conditional death proportion= $\tilde{q}_i = d_i / n_i$.

6. The conditional survival proportion $= \tilde{p}_i = 1 - \tilde{q}_i$

and finally the estimated survival function

$$\tilde{S}(\tau_0) = 1$$

$$\tilde{S}(\tau_1) = \tilde{S}(\tau_0) \cdot \tilde{p}_1$$

$$\tilde{S}(\tau_2) = \tilde{S}(\tau_1) \cdot \tilde{p}_2$$

etc.

(1) Interval	(2) start at risk Y_i	(3) no. of death d_i	(4) no. of cens. c_i	(5) mod. no. at risk n_i	(6) cond. death prob. $\tilde{q}_i$	(7) cond. surv. prob. $\tilde{p}_i$	(8) estimate of surv. fct. $\tilde{S}(\tau_i)$
0-60	*12*	*1*	*0*	*12*	*0.083*	*0.917*	**0.917**
60-120	*11*	*2*	*2*	*10*	*0.200*	*0.800*	**0.733**
120-180	*7*	*3*	*1*	*6.5*	*0.462*	*0.538*	**0.395**
180-240	*3*	*1*	*2*	*2*	*0.500*	*0.500*	**0.197**

Standard error of the life table estimate: Greenwood's formula

$$SE(\tilde{S}(\tau_i)) = \tilde{S}(\tau_i) \cdot \sqrt{\sum_{j=1}^{k} \frac{\tilde{q}_i}{\tilde{p}_i \cdot n_i}}$$

Example

At 60 days:

$$SE(\tilde{S}(60)) = \tilde{S}(60) \cdot \sqrt{\frac{\tilde{q}_1}{\tilde{p}_1 \cdot n_1}}$$

$$= 0.917 \cdot \sqrt{\frac{0.083}{0.917 \cdot 12}} = 0.080$$

At 120 days

$$SE(\tilde{S}(120)) = \tilde{S}(120) \cdot \sqrt{\frac{\tilde{q}_1}{\tilde{p}_1 \cdot n_1} + \frac{\tilde{q}_2}{\tilde{p}_2 \cdot n_2}}$$

$$= 0.917 \cdot \sqrt{\frac{0.083}{0.917 \cdot 12} + \frac{0.200}{0.800 \cdot 10}} = 0.132$$

etc.

USING STATA FOR THE ANALYSIS

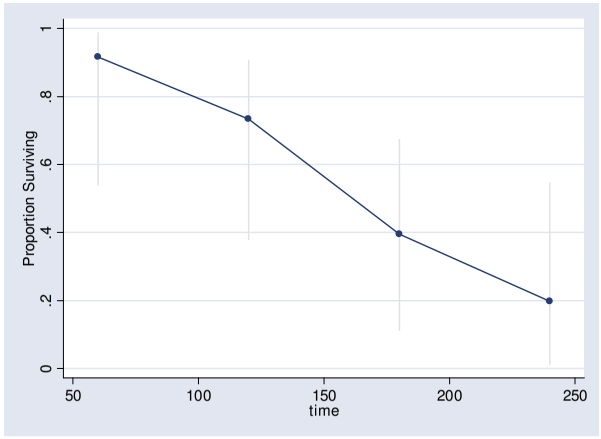
To get the life-table estimate of the survival function using Stata submit one of the commands

```
ltable time status, intervals(0,60,120,180,240)
ltable time status, intervals(0,60,120,180,240) graph
```

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

OUTPUT

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% CI.]	
0	60	12	1	0	0.9167	0.0798	0.5390	0.9878
60	120	11	2	2	0.7333	0.1324	0.3790	0.9056
120	180	7	3	1	0.3949	0.1601	0.1124	0.6737
180	240	3	1	2	0.1974	0.1609	0.0130	0.5455



COMPARING SURVIVAL IN TWO GROUPS

Standard non-parametric test can not 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 the 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.

Alternative tests*: Gehan's test* (also called Generalized Wilcoxon test or Breslow's test), *Tarone-Ware test*, *Peto-Prentice test* etc.

Several slightly different versions of these procedures exist. (differences are typically the result of applying different variance estimates).

Note:
We shall only consider unpaired two-sample problems here.

THE TWO-SAMPLE PROBLEM WITH CENSORED DATA

Problem:
Comparison of the survival in two groups of patients.

Data:
Two samples of life times, some of which may be censored.

Notation:

	Group 1	Group 2
Survival function	$S_1(t)$	$S_2(t)$
Hazard function	$\lambda_1(t)$	$\lambda_2(t)$

Null hypothesis: Same survival in the two groups.

$$S_1(t) = S_2(t) \qquad \text{for all } t$$

or, equivalently

$$\lambda_1(t) = \lambda_2(t) \qquad \text{for all } t$$

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

$$\lambda_1(t) \le \lambda_2(t) \qquad \text{for all } t$$

or

$$\lambda_1(t) \ge \lambda_2(t) \qquad \text{for all } t$$

The test statistics are particularly well suited (i.e. have high power), if the hazard rates are proportional:

$$\lambda_2(t) = \theta \cdot \lambda_1(t)$$

The parameter θ is called the ***hazard rate ratio*** or the ***mortality rate ratio***.

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

$$S_2(t) = (S_1(t))^{\theta}$$

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):

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

1) $\hat{\Lambda}_2(t)$ versus $\hat{\Lambda}_1(t)$

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

2) $\ln(\hat{\Lambda}_1(t))$ and $\ln(\hat{\Lambda}_2(t))$ against t in the same plot.

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

In both plots $\hat{\Lambda}_i(t)$ can be replaced by $-\ln(\hat{S}_i(t))$. In the second plot this leads to plots of $\ln(-\ln(\hat{S}_i(t)))$ against t.

THE LOG RANK TEST

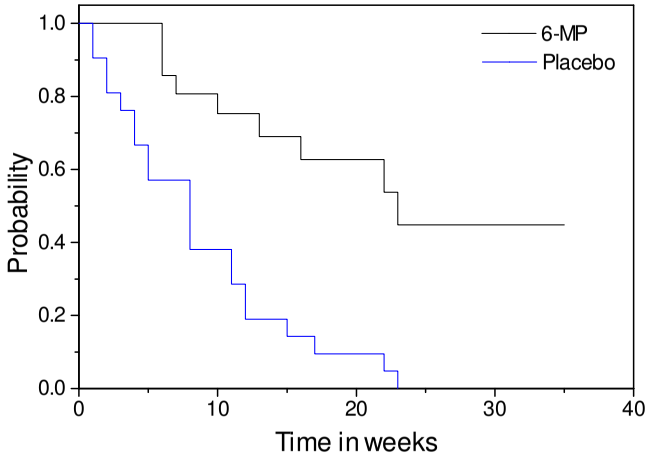
Example 1: A randomized, placebo-controlled trial to assess the effect of remission maintenance therapy with 6-MP on the duration of steroid-induced remission in acute leukemia.

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 trial (i.e. ignoring the paired design and sequential stopping rule).

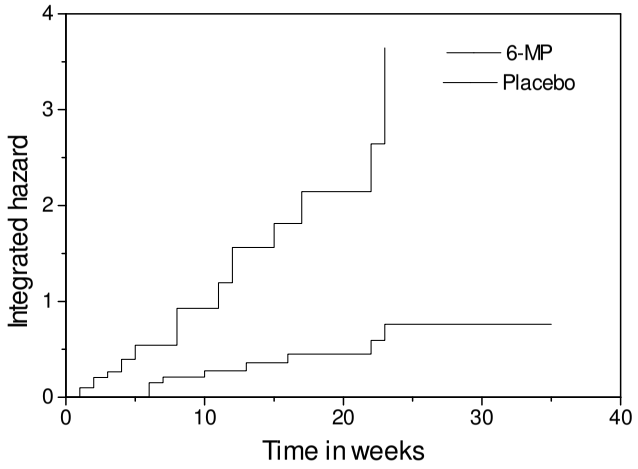
Data: Relapse-free survival in weeks.

6-MP Group			Placebo Group		
6	11+	25+	1	5	12
6	13	32+	1	8	15
6	16	32+	2	8	17
6+	17+	34+	2	8	22
7	19+	35+	3	8	23
9+	20+		4	11	
10	22		4	11	
10+	23		5	12	

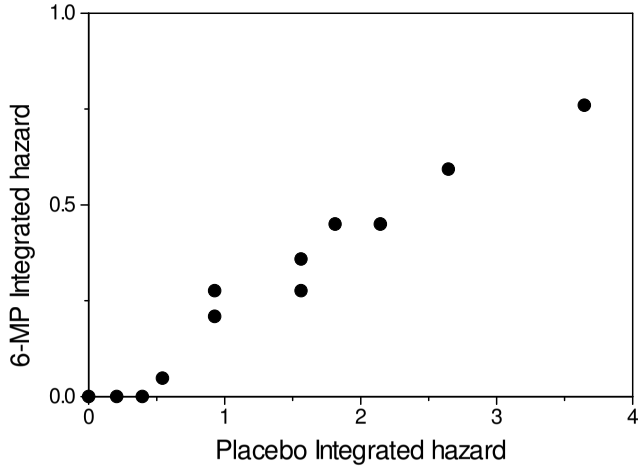
Estimated survival curves:

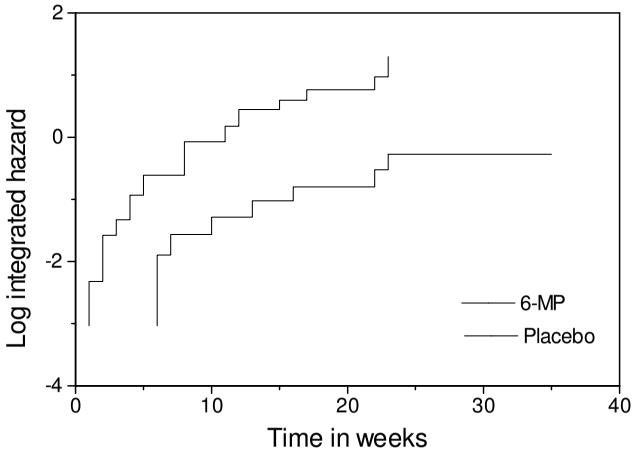


Estimated integrated hazards:



Diagnostic plots:





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 2×2 table is established:

Event time t_i	Treatment		
	6-MP	Placebo	Total
Relapse	m_{i1}	m_{i2}	m_i
No relapse			$Y(t_i) - m_i$
At risk at time t_i	$Y_1(t_i)$	$Y_2(t_i)$	$Y(t_i)$

The total m_i in the *Relapse* row is always 1, if no ties are present among uncensored observations.

At time t_i the proportion of patients at risk belonging to group 1 is $Y_1(t_i)/Y(t_i)$.

If the relapse rates are identical we will therefore at time t_i ***expect the same proportion of the relapses*** to occur in group 1, i.e. the expected number in group 1 is

$$m_i \cdot Y_1(t_i)/Y(t_i)$$

The data set in the example have 17 distinct relapse times, so this argument should be applied to 17 different 2×2 tables.

Time	Relapse	6-MP	Plac.	Total	Obs.	Expected
	Yes	0	2	2	0	$2 \cdot 21/42$
1	No	21	19	40		
	At risk	21	21	42		
	Yes	0	2	2	0	$2 \cdot 21/40$
2	No	21	17	38		
	At risk	21	19	40		
⋮		⋮	⋮	⋮	⋮	⋮
	Yes	1	0	1	1	$1 \cdot 15/23$
10	No	14	8	22		
	At risk	15	8	23		
⋮		⋮	⋮	⋮	⋮	⋮
	Yes	1	1	2	1	$2 \cdot 7/9$
22	No	6	1	7		
	At risk	7	2	9		
	Yes	1	1	2	1	$2 \cdot 6/7$
23	No	5	0	5		
	At risk	6	1	7		
Total					9	19.251

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 2×2 table in the same way as for the Mantel-Haenszel procedure.

For the first table

	6-MP	Placebo	Total
Relapse	0	2	2
No relapse	21	19	40
Total	21	21	42

the contribution to the variance becomes

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

$$\text{Variance} = V = 6.257$$

and the test statistic, ***the log rank test***, is obtained as

$$X^2_{lr} = \frac{(\text{Obs.} - \text{Exp.})^2}{\text{Variance}}$$

$$= \frac{(9 - 19.251)^2}{6.257}$$

$$= 16.79$$

On the null hypothesis of identical survival distributions the distribution of the test statistic is approximately a χ^2 distribution with 1 degree of freedom. Large values provide evidence against the null hypothesis.

A value of 16.79 is highly significant (p =0.00004).

The log rank test – in general

Consider two samples of possibly censored survival times. The sample sizes are denoted n_1 and n_2.

The observations in the combined data are sorted in ascending order. Ties among censored and uncensored values are resolved by taking uncensored values first. Let d denote the number of distinct values found among uncensored observations in the combined sample and denote these values by $t_1 < t_2 < \cdots < t_d$.

At time t_i the following 2×2 table can be established:

Event time t_i	Group		
	1	2	Total
Event	m_{i1}	m_{i2}	m_i
No event	$Y_1(t_i) - m_{i1}$	$Y_2(t_i) - m_{i2}$	$Y(t_i) - m_i$
At risk at time t_i	$Y_1(t_i)$	$Y_2(t_i)$	$Y(t_i)$

For each table compute

m_{i1} = the observed number of events in group 1 at time t_i.

e_{i1} = the expected number of events in group 1 at time t_i

$$= m_i \cdot \frac{Y_1(t_i)}{Y(t_i)}$$

v_i = the variance of m_{i1}

$$= \frac{m_i \cdot (Y(t_i) - m_i) \cdot Y_1(t_i) \cdot Y_2(t_i)}{Y(t_i)^2 \cdot (Y(t_i) - 1)}$$

The expected number and variance are computed on the hypothesis of identical hazard rates.

The contributions from each table are added together:

$O_1 = \sum_{i=1}^{d} m_{i1}$ Total ***observed number*** of events in group 1.

$E_1 = \sum_{i=1}^{d} e_{i1}$ Total ***expected number*** of events in group 1

$V = \sum_{i=1}^{d} v_i$ Total ***variance***.

The log rank test is finally obtained as

$$X_{lr}^2 = \frac{(O_1 - E_1)^2}{V}$$

On the null hypothesis X_{lr}^2 is approximately a χ^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. The standard terminology "observed" and "expected" is appealing, but not really justified. E_1 and E_2 ***are not expected values in the usual sense***, rather they are sums of conditional expectations. They represent our expectation based on the relative size of the two groups at the d different event times.

3. ***Some computer packages and text books use the name log rank test for a slightly different test statistic***, namely:

$$X^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

where the observed and expected numbers are those described above. ***The alternative version of the log rank test is (slightly) conservative***, since

$$X^2 < X_{lr}^2$$

is always satisfied. The p value from the alternative test is therefore too large. ***If the data contain no tied event times the discrepancy is minimal***, but the difference increases with the number of ties among the uncensored observations.

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*.

```
use e:\kurser\survival\aarhus2003\ex31.dta
stset time , failure(status==1)
```

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

```
sts test treatm
```

OUTPUT

```
. sts test treatm

        failure _d:  status == 1
  analysis time _t:  time


Log-rank test for equality of survivor functions

        |   Events         Events
treatm  |  observed       expected
--------+-------------------------
MP-6    |         9          19.25
Placebo |        21          10.75
--------+-------------------------
Total   |        30          30.00

             chi2(1) =      16.79
             Pr>chi2 =     0.0000
```

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

```
sts test treatm , noshow
```


Graphs: To get Kaplan-Meier survival curves for each group write

```
sts graph , by(treatm)

* to include confidence intervals write
sts graph , by(teatm) gwood

* to show each estimate in separate plot write
sts graph , by(treatm) separate
```

Graphs of Nelson-Aalen estimates: add option *na*

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

Tabular output:

```
sts list , by(treatm)
```

Format of listing similar to the one used to a single group

```
* to list estimates side-by-side
sts list , by(treatm) compare
```

Listing of Nelson-Aalen estimates: add option *na*

ALTERNATIVE NON-PARAMETRIC TEST

The log rank test is essentially a scaled version of the difference $O_1 - E_1$. Since

$$O_1 - E_1 = \sum_{i=1}^{d} m_{i1} - \sum_{i=1}^{d} e_{i1} = \sum_{i=1}^{d} (m_{i1} - e_{i1}) = \sum_{i=1}^{d} 1 \cdot (m_{i1} - e_{i1})$$

The contribution to the numerator from each 2×2 has the form $m_{i1} - e_{i1}$. In the log rank test these contributions are given equal weight.

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

$$X_w^2 = \frac{\sum_{i=1}^{d} w_i \cdot (m_{i1} - e_{i1})}{\sum_{i=1}^{d} w_i^2 \cdot v_i}$$

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

In STATA the following test statistics are available

Weight w_i	Name	STATA option
1	log rank	l (default)
$Y(t_i)$	Wilcoxon-Breslow-Gehan	w
$\hat{S}(t_i)$	Peto-Prentice	p
$\sqrt{Y(t_i)}$	Tarone-Ware	t
$\hat{S}(t_i)^p(1-\hat{S}(t_i))^q$	Fleming-Harrington	f(p q)

Gehan's test and Peto-Prentice's test both reduce to a Wilcoxon test if the data contain no censored observations.

Additional option: The likelihood ratio test (more on that later) in a proportional hazards model is computed with the option *cox*.

STATA output

```
. sts test treatm , w noshow

Wilcoxon (Breslow) test for equality of survivor
functions

         |   Events         Events        Sum of
treatm   |  observed       expected        ranks
---------+-------------------------------------
MP-6     |         9          19.25         -271
Placebo  |        21          10.75          271
---------+-------------------------------------
Total    |        30          30.00            0

              chi2(1) =      13.46
              Pr>chi2 =     0.0002

. sts test treatm , p noshow

Peto-Peto test for equality of survivor functions

         |   Events         Events        Sum of
treatm   |  observed       expected        ranks
---------+-------------------------------------
MP-6     |         9          19.25   -6.3622095
Placebo  |        21          10.75    6.3622095
---------+-------------------------------------
Total    |        30          30.00            0

              chi2(1) =      14.08
              Pr>chi2 =     0.0002
```

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 rates are approximately proportional this corresponds to estimating the constant of proportionality θ (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

"Treatment A reduces the mortality with 30% relative to treatment B",

i.e. $\lambda_A(t)/\lambda_B(t) = 0.70$, often described as the relative risk being 0.70, although it would be more appropriate to call the estimate a *relative rate.*

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 θ is obtained as

$$\theta^* = \frac{O_2 \cdot E_1}{E_2 \cdot O_1}$$

This estimate is known to be bias towards to 1, but the performance is adequate in the range ½ to 2.

Example (Freireich et al's data)

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

```
. sts test treatm , noshow

        |   Events         Events
treatm  |  observed       expected
--------+-------------------------
MP-6    |         9          19.25
Placebo |        21          10.75
--------+-------------------------
Total   |        30          30.00

             chi2(1) =      16.79
             Pr>chi2 =     0.0000
```

$$QaD\text{-estimate} = \frac{21 \cdot 19.25}{9 \cdot 10.75} = 4.18$$

```
. sts test treatm , c noshow

        |   Events         Events       Relative
treatm  |  observed       expected       hazard
--------+--------------------------------------
MP-6    |         9          19.25       0.5823
Placebo |        21          10.75       2.6338
--------+--------------------------------------
Total   |        30          30.00       1.0000

          LR chi2(1) =      15.21
             Pr>chi2 =     0.0001
```

$$\text{"Optimal" estimate} = \frac{2.6338}{0.5823} = 4.52$$

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

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: $\lambda_1(t) = \lambda_2(t) = \cdots = \lambda_K(t)$ for all t

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

```
. generate agegrp=age
. recode agegrp 0/39.999=1 40/49.999=2 50/59.999=3
  60/69.999=4 70/120=5
. sts test agegrp , noshow

        |   Events         Events
agegrp  |  observed       expected
--------+-------------------------
1       |         9          13.69
2       |        11          13.64
3       |        15          12.52
4       |        11          11.08
5       |        11           6.07
--------+-------------------------
Total   |        57          57.00

             chi2(4) =       6.75
             Pr>chi2 =     0.1496
```

THE *K* SAMPLE PROBLEM WITH ORDERED CATEGORIES: TEST FOR TREND.

Problem: Again, comparison of the survival in K ($K \geq 2$) 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:

$$\lambda_1(t) = \lambda_2(t) = \cdots = \lambda_K(t) \qquad \text{for all } t$$

In situations with a natural ordering of the groups it is often of special interest to investigate if ***the mortality increases (or decreases) with the natural ordering.***

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.

Solution: For each of the *K*-sample tests a corresponding ***test for trend*** has been developed. These test are particularly sensitive to monotone dependencies of the mortality rates on the ordering of the groups. Alternatively, regression models may be considered.

The log rank test for trend is a normalized version of the statistic

$$T = \sum_{j=1}^{d} x_j \cdot \left(O_j - E_j\right)$$

where 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).

The log rank test for trend:

$$X^2_{trend} = \frac{T^2}{V_T} = \frac{\left[\sum_{j=1}^{d} x_j \cdot \left(O_j - E_j\right)\right]^2}{V_T}$$

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

Note:
A linear transformations of the scores will not change the value of the test statistic, e.g.

Group	Scores *x*	Alt. scores
1	100	-1
2	145	0
3	190	1

The two set 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

```
. sts test agegrp , noshow trend

Log-rank test for equality of survivor functions

       |   Events         Events
agegrp |  observed       expected
-------+-------------------------
1      |          9          13.69
2      |         11          13.64
3      |         15          12.52
4      |         11          11.08
5      |         11           6.07
-------+-------------------------
Total  |         57          57.00

             chi2(4) =       6.75
             Pr>chi2 =     0.1496

Test for trend of survivor functions

             chi2(1) =       4.94
             Pr>chi2 =     0.0263
```


STRATIFIED ANALYSIS OF SURVIVAL DATA

Purpose:
A stratified analysis is a procedure to ***compare outcomes in different groups*** and at the same time ***correct for*** (or reduce) ***the effect of other factors*** which are related to the outcome but not identically distributed in the different groups (so-called ***confounding factors***).

An example:
Suppose that for a particular disease male patients are known to have a worse prognosis than female patients. If the proportion of males differ markedly between two treatment groups a comparison based on a simple two sample test will be misleading: The result of a statistical test will not only reflect a possible difference in efficacy, but also the difference in the proportion of patients with poorer prognosis.

A stratified analysis can here be used to compare the two treatments and at the same time adjust for the effect of the difference in the sex distribution.

The basic idea:
In a stratified analysis the comparability is improved by dividing the patients up into smaller and more homogeneous groups. The basic comparison is made within such groups (or strata), and the discrepancy found is accumulated across all strata. The test statistic is derived from these accumulated statistics.

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 life times, 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 O_A and O_B, the corresponding ***expected numbers*** E_A and E_B and the ***variance*** V.. 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:

$$X_S^2 = \frac{\left(\sum_{\text{strata}} O_A - \sum_{\text{strata}} E_A\right)^2}{\sum_{\text{strata}} V}$$

Stratum 1: Age < 30.

Treatment	Observed	Expected	Variance
A	O_{A1}	E_{A1}	V_1
B	O_{B1}	E_{B1}	

Stratum 2: 30 ≤ Age < 50.

Treatment	Observed	Expected	Variance
A	O_{A2}	E_{A2}	V_2
B	O_{B2}	E_{B2}	

Stratum 3: 50 ≤ Age < 70.

Treatment	Observed	Expected	Variance
A	O_{A3}	E_{A3}	V_3
B	O_{B3}	E_{B3}	

Stratum 4: 70 ≤ Age.

Treatment	Observed	Expected	Variance
A	O_{A4}	E_{A4}	V_4
B	O_{B4}	E_{B4}	

Summed across strata:

Treatment	Observed	Expected	Variance
A	$\sum_{k=1}^{4} O_{Ak}$	$\sum_{k=1}^{4} E_{Ak}$	$\sum_{k=1}^{4} V_k$
B	$\sum_{k=1}^{4} O_{Bk}$	$\sum_{k=1}^{4} E_{Bk}$	

The stratified log rank test is based on these statistics

Comments:

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 may alternatively be used 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 variation in the treatment difference between strata (treatment×strata interaction). The stratified analysis implicitly assumes that a treatment difference has roughly the same size in all 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 derived in a similar fashion. 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 from these sums.

STRATIFIED ANALYSIS USING STATA

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

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

If output should include separate comparisons for each strata add the option `detail`. Apparently, only one stratum variable is allow when `detail` is specified. This problem (bug?) can be solved in the following way:

```
egen agesex=group(agecat sex)
sts test expo , strata(agesex) detail
```

A Peto-Prentice test for trend in age (*agecat*) stratified by *sex* with detailed output

```
sts test agecat , ///
peto strata(sex) trend detail
```