power logrank, cluster — Power analysis for the log-rank test, CRD

Description

power logrank, cluster computes group-specific numbers of clusters, group-specific cluster sizes, power, or the hazard ratio for the log-rank test comparing survivor functions in two groups in a cluster randomized design (CRD). Without censoring, the survival input parameter is the hazard ratio; otherwise, the survival input parameters are the control-group and experimental-group survival probabilities.

power logrank, cluster computes group-specific numbers of clusters given cluster sizes, power, and survival parameters. It also computes group-specific cluster sizes given numbers of clusters, power, and survival parameters. Alternatively, it computes power given numbers of clusters, cluster sizes, and survival parameters, or it computes the hazard ratio given numbers of clusters, cluster sizes, power, and, in the presence of censoring, the control-group survival probability. See [PSS-2] power logrank for a general discussion of power and sample-size analysis for the log-rank test. Also see [PSS-2] power for a general introduction to the power command using hypothesis tests.

Quick start

Numbers of clusters for uncensored design with alternative hazard ratio $\Delta_a = 0.76$ for the log-rank test of $H_0$: $\Delta = 1$ versus $H_a$: $\Delta \neq 1$ given a cluster size of 5 in both groups and using default intraclass correlation of 0.5, power of 0.8, and significance level $\alpha = 0.05$

```
power logrank, m1(5) m2(5) hratio(0.76)
```

As above, but use intraclass correlation of 0.4

```
power logrank, m1(5) m2(5) hratio(0.76) rho(0.4)
```

Assume cluster sizes vary with an average cluster size of 5 and a coefficient of variation of 0.6

```
power logrank, m1(5) m2(5) hratio(0.76) cvcluster(0.6)
```

Group-specific numbers of clusters using ratio of experimental clusters to control clusters of 0.5

```
power logrank, m1(5) m2(5) hratio(0.76) kratio(0.5)
```

Cluster sizes for censored design with control- and experimental-group survival probabilities of 0.3 and 0.4 for the log-rank test of $H_0$: $\Delta = 1$ versus $H_a$: $\Delta \neq 1$ with 200 clusters in both groups using default intraclass correlation of 0.5, power of 0.8, and significance level $\alpha = 0.05$

```
power logrank 0.3 0.4, k1(200) k2(200)
```

Power for 50 clusters of size 5 in the control group and 50, 100, 150, or 200 clusters of size 5 in the experimental group with results shown in a power plot

```
power logrank 0.3 0.4, k1(50) k2(50(50)200) m1(5) m2(5) graph
```

Hazard ratio and experimental-group survival probability with power of 0.8

```
power logrank 0.3, k1(50) k2(50) m1(5) m2(5) power(0.8)
```
Menu
Statistics > Power, precision, and sample size

Syntax

Compute numbers of clusters

```
power logrank [surv1 [surv2]], {mspec | nspec | cluster} [options]
```

Compute cluster sizes

```
power logrank [surv1 [surv2]], kspec [options]
```

Compute power

```
power logrank [surv1 [surv2]], kspec {mspec | nspec} [options]
```

Compute effect size

```
power logrank [surv1], kspec {mspec | nspec} power(numlist) [options]
```

where `surv1` is the survival probability in the control (reference) group at the end of the study $t^*$ and `surv2` is the survival probability in the experimental (comparison) group at the end of the study $t^*$. `surv1` and `surv2` may each be specified either as one number or as a list of values in parentheses (see [U] 11.1.8 numlist).

`mspec` is one of

- `m1()` `m2()`
- `m1() [mratio()]`
- `m2() [mratio()]`

`nspec` is one of

- `n1()` `n2()`
- `n1() [nratio()]`
- `n2() [nratio()]`

`kspec` is one of

- `k1()` `k2()`
- `k1() [kratio()]`
- `k2() [kratio()]`
**power logrank, cluster** — Power analysis for the log-rank test, CRD

<table>
<thead>
<tr>
<th>options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
<td></td>
</tr>
<tr>
<td><strong>cluster</strong></td>
<td>perform computations for a CRD; implied by k1(), k2(), m1(), or m2()</td>
</tr>
<tr>
<td><em>alpha(numlist)</em></td>
<td>significance level; default is alpha(0.05)</td>
</tr>
<tr>
<td><em>power(numlist)</em></td>
<td>power; default is power(0.8)</td>
</tr>
<tr>
<td><em>beta(numlist)</em></td>
<td>probability of type II error; default is beta(0.2)</td>
</tr>
<tr>
<td><em>k1(numlist)</em></td>
<td>number of clusters in the control group</td>
</tr>
<tr>
<td><em>k2(numlist)</em></td>
<td>number of clusters in the experimental group</td>
</tr>
<tr>
<td><em>kratio(numlist)</em></td>
<td>cluster ratio, K2/K1; default is kratio(1)</td>
</tr>
<tr>
<td><em>m1(numlist)</em></td>
<td>cluster size of the control group</td>
</tr>
<tr>
<td><em>m2(numlist)</em></td>
<td>cluster size of the experimental group</td>
</tr>
<tr>
<td><em>mratio(numlist)</em></td>
<td>cluster-size ratio, M2/M1; default is mratio(1)</td>
</tr>
<tr>
<td><em>n1(numlist)</em></td>
<td>sample size of the control group</td>
</tr>
<tr>
<td><em>n2(numlist)</em></td>
<td>sample size of the experimental group</td>
</tr>
<tr>
<td><em>nratio(numlist)</em></td>
<td>sample-size ratio, N2/N1; default is nratio(1)</td>
</tr>
<tr>
<td><strong>nfractional</strong></td>
<td>allow fractional numbers of clusters, cluster sizes, and sample sizes</td>
</tr>
<tr>
<td><em>hratio(numlist)</em></td>
<td>hazard ratio of the experimental to the control group; default is hratio(0.5)</td>
</tr>
<tr>
<td><em>lnhratio(numlist)</em></td>
<td>log hazard-ratio of the experimental to the control group</td>
</tr>
<tr>
<td><em>rho(numlist)</em></td>
<td>intraclass correlation; default is rho(0.5)</td>
</tr>
<tr>
<td><em>cvcluster(numlist)</em></td>
<td>coefficient of variation for cluster sizes</td>
</tr>
<tr>
<td>**direction(lower</td>
<td>upper)**</td>
</tr>
<tr>
<td><strong>onesided</strong></td>
<td>one-sided test; default is two sided</td>
</tr>
<tr>
<td><strong>parallel</strong></td>
<td>treat number lists in starred options or in command arguments as parallel when multiple values per option or argument are specified (do not enumerate all possible combinations of values)</td>
</tr>
</tbody>
</table>

**Table**

* no table *(tablespec)* | suppress table or display results as a table; see [PSS-2] power, table |
* saving(filename [, replace]) | save the table data to filename; use replace to overwrite existing filename |

**Graph**

* graph *(graphopts)* | graph results; see [PSS-2] power, graph |
**power logrank, cluster** — Power analysis for the log-rank test, CRD

**Iteration**

- `init(#)`: initial value for hazard ratio
- `iterate(#)`: maximum number of iterations; default is `iterate(500)`
- `tolerance(#)`: parameter tolerance; default is `tolerance(1e-12)`
- `ftolerance(#)`: function tolerance; default is `ftolerance(1e-12)`
- `[no] log`: suppress or display iteration log
- `[no] dots`: suppress or display iterations as dots
- `notitle`: suppress the title

*Specifying a list of values in at least two starred options, or at least two command arguments, or at least one starred option and one argument results in computations for all possible combinations of the values; see [U] 11.1.8 numlist. Also see the parallel option.

`notitle` does not appear in the dialog box.

where `tablespec` is

```
  column[ :label ] [ column[ :label ] ... ] [ , tableopts ]
```

`column` is one of the columns defined below, and `label` is a column label (may contain quotes and compound quotes).

<table>
<thead>
<tr>
<th>column</th>
<th>Description</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>significance level</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>power</td>
<td>power</td>
<td>$1 - \beta$</td>
</tr>
<tr>
<td>beta</td>
<td>type II error probability</td>
<td>$\beta$</td>
</tr>
<tr>
<td>K1</td>
<td>number of clusters in the control group</td>
<td>$K_1$</td>
</tr>
<tr>
<td>K2</td>
<td>number of clusters in the experimental group</td>
<td>$K_2$</td>
</tr>
<tr>
<td>kratio</td>
<td>ratio of numbers of clusters, experimental to control</td>
<td>$K_2/K_1$</td>
</tr>
<tr>
<td>M1</td>
<td>cluster size of the control group</td>
<td>$M_1$</td>
</tr>
<tr>
<td>M2</td>
<td>cluster size of the experimental group</td>
<td>$M_2$</td>
</tr>
<tr>
<td>mratio</td>
<td>ratio of cluster sizes, experimental to control</td>
<td>$M_2/M_1$</td>
</tr>
<tr>
<td>N</td>
<td>total number of observations</td>
<td>$N$</td>
</tr>
<tr>
<td>N1</td>
<td>number of observations in the control group</td>
<td>$N_1$</td>
</tr>
<tr>
<td>N2</td>
<td>number of observations in the experimental group</td>
<td>$N_2$</td>
</tr>
<tr>
<td>nratio</td>
<td>ratio of sample sizes, experimental to control</td>
<td>$N_2/N_1$</td>
</tr>
<tr>
<td>delta</td>
<td>effect size</td>
<td>$\delta$</td>
</tr>
<tr>
<td>E</td>
<td>total number of events (failures)</td>
<td>$E$</td>
</tr>
<tr>
<td>hratio</td>
<td>hazard ratio</td>
<td>$\Delta$</td>
</tr>
<tr>
<td>lnhratio</td>
<td>log hazard-ratio</td>
<td>$\ln(\Delta)$</td>
</tr>
<tr>
<td>s1</td>
<td>survival probability in the control group</td>
<td>$S_1(T)$</td>
</tr>
<tr>
<td>s2</td>
<td>survival probability in the experimental group</td>
<td>$S_2(T)$</td>
</tr>
<tr>
<td>Pr_E</td>
<td>overall probability of an event (failure)</td>
<td>$p_E$</td>
</tr>
<tr>
<td>rho</td>
<td>intraclass correlation</td>
<td>$\rho$</td>
</tr>
<tr>
<td>CV_cluster</td>
<td>coefficient of variation for cluster sizes</td>
<td>$CV_{cl}$</td>
</tr>
<tr>
<td>target</td>
<td>target parameter; hratio</td>
<td></td>
</tr>
<tr>
<td>_all</td>
<td>display all supported columns</td>
<td></td>
</tr>
</tbody>
</table>
Column beta is shown in the default table in place of column power if specified.
Column N is shown in the table if specified.
Columns N1 and N2 are shown in the default table if n1() or n2() is specified.
Column lnhratio is shown in the default table in place of column hratio if specified.
Columns s1 and s2 are available only when specified.
Columns nratio and CV_cluster are shown in the default table if specified.

Options

```
Main

cluster specifies that computations should be performed for a CRD. This option is implied when the k1(), k2(), m1(), or m2() option is specified. cluster is required to compute the numbers of clusters when nspec is used to specify sample sizes instead of mspec for cluster sizes.

alpha(), power(), beta(); see [PSS-2] power.

k1(numlist) specifies the number of clusters in the control group.
k2(numlist) specifies the number of clusters in the experimental group.
kratio(numlist) specifies the ratio of the numbers of clusters of the experimental group relative to the control group, K2/K1. The default is kratio(1), meaning equal numbers of clusters in the two groups.
m1(numlist) specifies the cluster size of the control group. m1() may contain noninteger values.
m2(numlist) specifies the cluster size of the experimental group. m2() may contain noninteger values.
mratio(numlist) specifies the ratio of cluster sizes of the experimental group relative to the control group, M2/M1. The default is mratio(1), meaning equal cluster sizes in the two groups.
n1(), n2(), nratio(); see [PSS-2] power.
nfractional; see [PSS-2] power. The nfractional option displays fractional (without rounding) values of the numbers of clusters, cluster sizes, and sample sizes.
hratio(), lnhratio(); see [PSS-2] power logrank.
rho(numlist) specifies the intraclass correlation. The default is rho(0.5).
cvcluster(numlist) specifies the coefficient of variation for cluster sizes. This option is used with varying cluster sizes.
direction(), onesided, parallel; see [PSS-2] power. direction(lower) is the default.
```

```
Table

table, table(), notable; see [PSS-2] power, table.
saving(); see [PSS-2] power.
```

```
Graph

graph, graph(); see [PSS-2] power, graph. Also see the column table for a list of symbols used by the graphs.
```
The following option is available with power logrank, cluster but is not shown in the dialog box: notitle; see [PSS-2] power.

Remarks and examples

Remarks are presented under the following headings:

- Using power logrank, cluster
- Computing numbers of clusters
- Computing cluster sizes
- Computing power
- Computing effect size
- Compare two survivor functions with clustered data

power logrank, cluster requests that computations for the power logrank command be done for a CRD. In a CRD, groups of subjects or clusters are randomized instead of individual subjects, so the sample size is determined by the numbers of clusters and the cluster sizes. The sample-size determination thus consists of the determination of the numbers of clusters given cluster sizes or the determination of cluster sizes given the numbers of clusters. For a general discussion of using power logrank, see [PSS-2] power logrank. The discussion below is specific to the CRD.

Using power logrank, cluster

If you specify the cluster option, include k1() or k2() to specify the number of clusters or include m1() or m2() to specify the cluster size, the power logrank command will perform computations for the log-rank test in a CRD. The computations for a CRD are based on the Freedman method; see Introduction in [PSS-2] power logrank for details.

All computations are performed for a two-sided hypothesis test where, by default, the significance level is set to 0.05. You may change the significance level by specifying the alpha() option. You can specify the onesided option to request a one-sided test. By default, all computations assume a balanced or equal-allocation design, meaning equal numbers of clusters and cluster sizes in both groups; see [PSS-4] Unbalanced designs for a description of how to specify an unbalanced design.

To compute the number of clusters in both groups, you must provide cluster sizes for both groups. There are multiple ways to supply cluster sizes, but the most common is to specify the cluster size of the control group in the m1() option and the cluster size of the experimental group in the m2() option. See mspec and nspec under Syntax for other specifications. When nspec is specified, the cluster option is also required to request that power logrank perform computations for a CRD. The number of clusters is assumed to be equal in the two groups, but you can change this by specifying the ratio of the numbers of clusters in the experimental to the control group in the kratio() option. Other parameters are specified as described in Using power logrank in [PSS-2] power logrank.

To compute the cluster sizes in both groups, you must provide the numbers of clusters in both groups. There are several ways to supply the numbers of clusters; see kspec under Syntax. The most common is to specify the numbers of clusters in the control group and the experimental group in the k1() and k2() options, respectively. Equal cluster sizes are assumed in the two groups, but you can change this by specifying the ratio of the cluster sizes in the experimental to that of the control group in the mratio() option. Other parameters are specified as described in Using power logrank in [PSS-2] power logrank.
The power and effect-size determination is the same as described in Using power logrank in [PSS-2] power logrank, but the sample-size information is supplied as the numbers of clusters \( kspec \) and either cluster sizes using \( mspec \) or, less commonly, sample sizes using \( nspec \).

All computations assume an intraclass correlation of 0.5. You can change this by specifying the \( \rho() \) option. Also, all clusters are assumed to be of the same size unless the coefficient of variation for cluster sizes is specified in the \( cvcluster() \) option.

By default, the computed numbers of clusters, cluster sizes, and sample sizes are rounded up. However, you can specify the \( nfractional \) option to see the corresponding fractional values; see Fractional sample sizes in [PSS-4] Unbalanced designs for an example. If the \( cvcluster() \) option is specified when computing cluster sizes, then cluster sizes represent average cluster sizes and are thus not rounded. When sample sizes are specified using \( nspec \), fractional cluster sizes may be reported to accommodate the specified numbers of clusters and sample sizes.

Some of power logrank, cluster’s computations require iteration; see Methods and formulas for details and [PSS-2] power for the descriptions of options that control the iteration procedure.

### Computing numbers of clusters

To compute the numbers of clusters in each group, you must either provide the cluster size for each group using \( mspec \) or specify the \( cluster \) option and provide the sample sizes of both groups using \( nspec \). The most common method is to use \( mspec \) of \( m1() \) and \( m2() \). A hazard ratio of 0.5 is assumed but may be changed by specifying the \( hratio() \) option. If there is censoring, the control-group survival probability \( surv_1 \) must be specified, and the experimental-group survival probability \( surv_2 \) may be specified instead of the hazard ratio.

> **Example 1: Numbers of clusters for the log-rank test with no censoring in a CRD**

Consider an example from Xie and Waksman (2003). A researcher would like to study the time it takes diabetic foot ulcers to heal completely (the event of interest). Patients enrolled in the study may have multiple ulcers, and all existing ulcers will be treated. Suppose that at the end of the study, all ulcers are expected to be healed. Patients, the clustering unit, are randomly assigned equally to the control group with the usual treatment and the experimental group with the new treatment. The researcher is interested in detecting a hazard ratio, experimental group to control group, of 1.79 for healing. For the planned study, the researcher assumes about 3 ulcers per patient in each group with an intraclass correlation of 0.3.

To compute the numbers of patients in each group required to detect a hazard ratio of 1.79 with 80% power using a 5%-level two-sided test, we type
. power logrank, hratio(1.79) m1(3) m2(3) rho(0.3)

Estimated numbers of clusters for two-sample comparison of survivor functions
Cluster randomized design, log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1

Study parameters:
  alpha = 0.0500
  power = 0.8000
  delta = 1.7900 (hazard ratio)
  hratio = 1.7900

Cluster design:
  M1 = 3
  M2 = 3
  rho = 0.3000

Number of events and censoring:
  E = 157
  Pr_E = 1.0000

Estimated numbers of clusters and sample sizes:
  K1 = 27
  K2 = 27
  N1 = 81
  N2 = 81

We find that 27 patients per group are required to detect a hazard ratio of 1.79 with 80% power using a 5%-level two-sided test.

Example 2: Numbers of clusters for the log-rank test with censoring in a CRD

Unlike example 1, suppose that 70% of the ulcers in the control group and 50% in the experimental group are not healed at the end of the study. Because there is censoring, we specify the survival probabilities of 0.7 and 0.5 as command arguments.
. power logrank 0.7 0.5, m1(3) m2(3) rho(0.3)

Estimated numbers of clusters for two-sample comparison of survivor functions
Cluster randomized design, log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1

Study parameters:
  alpha = 0.0500
  power = 0.8000
  delta = 1.9434 (hazard ratio)
  hratio = 1.9434

Cluster design:
  M1 = 3
  M2 = 3
  rho = 0.3000

Number of events and censoring:
  E = 123
  s1 = 0.7000
  s2 = 0.5000
  Pr_E = 0.4000

Estimated numbers of clusters and sample sizes:
  K1 = 51
  K2 = 51
  N1 = 153
  N2 = 153

We find that 51 patients per group are required to detect a decrease in the survival rate or an increase in the healing rate of ulcers from 0.7 to 0.5 with 80% power using a 5%-level two-sided test.

Example 3: Numbers of clusters for the log-rank test in a CRD, varying cluster sizes

Continuing with example 2, we now want to account for the fact that the numbers of ulcers may vary among patients. We assume a coefficient of variation of 0.4. To compute the numbers of clusters when cluster sizes vary, we specify the coefficient of variation of the numbers of ulcers of 0.4 in the cvcluster() option.
Estimated numbers of clusters for two-sample comparison of survivor functions
Cluster randomized design, log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1

Study parameters:

\[
\begin{align*}
\text{alpha} & = 0.0500 \\
\text{power} & = 0.8000 \\
\text{delta} & = 1.9434 \quad \text{(hazard ratio)} \\
\text{hratio} & = 1.9434 \\
\end{align*}
\]

Cluster design:

\[
\begin{align*}
\text{Average M1} & = 3.0000 \\
\text{Average M2} & = 3.0000 \\
\text{rho} & = 0.3000 \\
\text{CV_cl} & = 0.4000 \\
\end{align*}
\]

Number of events and censoring:

\[
\begin{align*}
\text{E} & = 134 \\
\text{s1} & = 0.7000 \\
\text{s2} & = 0.5000 \\
\text{Pr}_E & = 0.4000 \\
\end{align*}
\]

Estimated numbers of clusters and sample sizes:

\[
\begin{align*}
\text{K1} & = 56 \\
\text{K2} & = 56 \\
\text{N1} & = 168 \\
\text{N2} & = 168 \\
\end{align*}
\]

The required number of patients in each group is 56, which is larger than the required number of patients of 51 in example 2. When the numbers of ulcers vary among patients, we need more patients to achieve the same power.

Computing cluster sizes

To compute cluster sizes in both groups, you must provide the numbers of clusters in both groups by using `kspec`. The most common method is to specify the numbers of clusters in the control and experimental groups in the `k1()` and `k2()` options, respectively. As for the determination of the numbers of clusters, a hazard ratio of 0.5 is assumed but may be changed by specifying the `hratio()` option. If there is censoring, the control-group survival probability `surv1` must be specified, and the experimental-group survival probability `surv2` may be specified instead of the hazard ratio.

Example 4: Cluster sizes for the log-rank test in a CRD

Continuing with example 2, suppose that we are designing a new study and planning to recruit 100 patients, with 50 patients in each group. Given other study parameters from example 2, we compute the number of ulcers we would expect to see from each patient to achieve the power of 80% by specifying 50 clusters in the `k1()` and `k2()` options.
. power logrank 0.7 0.5, k1(50) k2(50) rho(0.3)

Estimated cluster sizes for two-sample comparison of survivor functions
Cluster randomized design, log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1

Study parameters:
- alpha = 0.0500
- power = 0.8000
- delta = 1.9434 (hazard ratio)
- hratio = 1.9434

Cluster design:
- K1 = 50
- K2 = 50
- rho = 0.3000

Number of events and censoring:
- E = 77
- s1 = 0.7000
- s2 = 0.5000
- Pr_E = 0.4000

Estimated cluster sizes and sample sizes:
- M1 = 4
- M2 = 4
- N1 = 200
- N2 = 200

To achieve the desired power of 80%, with 100 patients, we will need to observe 4 ulcers per patient.

Computing power

To compute power in a CRD, you supply the sample-size information as the numbers of clusters by using \(kspec\) along with either the cluster sizes by using \(mspec\) or, less commonly, the sample sizes by using \(nspec\). The most common method is to specify the \(k1()\), \(k2()\), \(m1()\), and \(m2()\) options. A hazard ratio of 0.5 is assumed but may be changed by specifying the \(hratio()\) option. If there is censoring, the control-group survival probability \(surv_1\) must be specified, and the experimental-group survival probability \(surv_2\) may be specified instead of the hazard ratio.

Example 5: Power for the log-rank test in a CRD

Continuing with example 2, suppose that we can recruit 100 patients (50 patients per group) and observe 3 ulcers per patient and we want to compute power for this design. Given other study parameters from example 2, we compute power by specifying 50 clusters in the \(k1()\) and \(k2()\) options and 3 as the cluster size in the \(m1()\) and \(m2()\) options.
Estimated power for two-sample comparison of survivor functions
Cluster randomized design, log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR ≠ 1

Study parameters:

- \( \alpha = 0.0500 \)
- \( \delta = 1.9434 \) (hazard ratio)
- \( h_{ratio} = 1.9434 \)

Cluster design:

- \( K_1 = 50 \)
- \( K_2 = 50 \)
- \( M_1 = 3 \)
- \( M_2 = 3 \)
- \( N_1 = 150 \)
- \( N_2 = 150 \)
- \( \rho = 0.3000 \)

Number of events and censoring:

- \( E = 120 \)
- \( s_1 = 0.7000 \)
- \( s_2 = 0.5000 \)
- \( Pr_E = 0.4000 \)

Estimated power:

- \( power = 0.7927 \)

The computed power is about 79%.

Example 6: Multiple values of study parameters

To investigate the effect of the number of clusters in the experimental group on power, we can specify a list of numbers of clusters in the \( k_2() \) option:

```
.power logrank 0.7 0.5, k1(50) k2(10(20)90) m1(3) m2(3) rho(0.3) table(power K2)
```

Estimated power for two-sample comparison of survivor functions
Cluster randomized design, log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR ≠ 1

<table>
<thead>
<tr>
<th>power</th>
<th>K2</th>
</tr>
</thead>
<tbody>
<tr>
<td>.4603</td>
<td>10</td>
</tr>
<tr>
<td>.7157</td>
<td>30</td>
</tr>
<tr>
<td>.7927</td>
<td>50</td>
</tr>
<tr>
<td>.8276</td>
<td>70</td>
</tr>
<tr>
<td>.8472</td>
<td>90</td>
</tr>
</tbody>
</table>

In this example, we also specified the `table(power K2)` option to list only the two columns that vary. As expected, as the number of clusters increases, the power tends to get closer to 1.

For multiple values of parameters, the results are automatically displayed in a table, as we see above. For more examples of tables, see [PSS-2] `power, table`. If you wish to produce a power plot, see [PSS-2] `power, graph`. 
Computing effect size

Effect size \( \delta \) for the log-rank test is the hazard ratio. To compute effect size in a CRD, you supply the sample-size information as the numbers of clusters by using \textit{kspec} along with either the cluster sizes by using \textit{mspec} or, less commonly, the sample sizes by using \textit{nspec}. The most common method is to specify the \texttt{k1()}, \texttt{k2()}, \texttt{m1()}, and \texttt{m2()} options. In addition, power and, in the presence of censoring, control-group survival probability must be specified. You must also decide on the direction of the effect, which is specified in the \texttt{direction()} option. For the default, lower, meaning \( \Delta < 1 \), \texttt{power logrank, cluster} uses \texttt{direction(lower)}. For upper, meaning \( \Delta > 1 \), specify \texttt{direction(upper)}.

\section*{Example 7: Effect size for the log-rank test in a CRD}

Continuing with example 5, we may be interested in finding the minimum value of the hazard ratio that can be detected with a sample of 50 patients per group, with about 3 ulcers per patient, and with 80\% power. To compute this, we specify the control-group survival probability of 0.7 as the command argument and required options \texttt{k1(50), k2(50), m1(3), m2(3), and power(0.8)}. We continue to use \texttt{rho(0.3)}. In this example, we want to detect an effect in the upper direction. That is, we believe the rate of healing in the experimental group is greater than in the control group, so we expect a hazard ratio greater than 1. Therefore, we need to specify the \texttt{direction(upper)} option.

\begin{verbatim}
. power logrank 0.7, k1(50) k2(50) m1(3) m2(3) power(0.8) rho(0.3)
> direction(upper)
Performing iteration ...
Estimated hazard ratio for two-sample comparison of survivor functions
Cluster randomized design, log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1; HR > 1
Study parameters:
  alpha =  0.0500
  power =  0.8000
Cluster design:
  K1 =  50
  K2 =  50
  M1 =   3
  M2 =   3
  N1 =  150
  N2 =  150
  rho =  0.3000
Number of events and censoring:
  E =  121
  s1 =  0.7000
  s2 =  0.4980
  Pr_E =  0.4010
Estimated effect size and hazard ratio:
  delta =  1.9546 (hazard ratio)
  hratio =  1.9546
\end{verbatim}

The minimum detectable value of the hazard ratio is about 1.95.
Compare two survivor functions with clustered data

To compare two survivor functions nonparametrically, we can use the `sts test` command to perform, for example, the log-rank test. `sts test`, however, assumes independent data. We can compare the survivor functions of the two groups semiparametrically while accounting for correlated observations by using a Cox proportional hazards model with clustered standard errors. The Cox model assumes proportional hazards, so testing the equality of two survivor functions reduces to testing that the ratio of the corresponding hazard functions, the hazard ratio, is equal to one. So we can include an identifier of the two groups in the Cox model as the covariate and test the corresponding hazard ratio against one or, equivalently, the log hazard-ratio (coefficient) against zero.

In this section, we briefly demonstrate the `stcox` command with clustered standard errors to compare two groups with survival-time clustered data.

Example 8: Comparing two survivor functions with clustered data using stcox

Consider `eartubes.dta` containing durations of ventilating tubes from Le and Lindgren (1996). To prevent frequent ear infections in infants and young children, a surgical procedure of inserting ventilating tubes in the ears is often recommended. In this study, children with ventilating tubes in both ears were assigned to either a treatment group (receiving a steroid treatment postsurgery) or a control group (receiving no postsurgery intervention). We want to compare the duration of ventilating tubes for children in the two groups. Because durations of ventilating tubes within a child, from the left and right ears, will be correlated, we cluster on the child identifier, `childid`, in our analysis.

We first declare our survival-time data with the `stset` command and then use the `stcox` command with the `vce(cluster childid)` option to fit a Cox model with a treatment variable identifying the two groups as the covariate.

```stata
use https://www.stata-press.com/data/r16/eartubes
(Ventilating tubes life data (Le and Lindgren (1996)))
.stset time, failure(status)
    failure event: status != 0 & status < .
ob. time interval: (0, time]
exit on or before: failure

156  total observations
0    exclusions

156  observations remaining, representing
144  failures in single-record/single-failure data
1,373.8 total analysis time at risk and under observation
    at risk from t = 0
    earliest observed entry t = 0
    last observed exit t = 33
```
The estimated hazard ratio is 0.7, and the test of log hazard-ratio (or coefficient) of treatment against zero has a $p$-value = 0.059. We do not have sufficient statistical evidence to conclude that the durations of ventilating tubes of the two groups are different, at least at the 5% significance level.

Suppose that we want to use the results of this study to design a new study. We want to determine how many children we need in the study. To compute the required number of children or clusters, we need the number of observations per cluster in each group and the estimates of the hazard ratio, of the intraclass correlation, and of the survival rate in the control group at the end of the study.

We can use the estimate of 0.7 for the hazard ratio from \texttt{stcox}. Suppose that we are planning to stop our new study after 12 months. We can use the \texttt{sts list} command to estimate the control-group survival rate at 12 months based on these data,

```
. sts list if treatment==0, at(12 12)
   failure _d: status
   analysis time _t: time
```

<table>
<thead>
<tr>
<th>Time</th>
<th>Beg. Total</th>
<th>Fail</th>
<th>Survivor Function</th>
<th>Std. Error</th>
<th>[95% Conf. Int.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>15</td>
<td>56</td>
<td>0.2035</td>
<td>0.0484</td>
<td>0.1184 0.3049</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>0</td>
<td>0.2035</td>
<td>0.0484</td>
<td>0.1184 0.3049</td>
</tr>
</tbody>
</table>

Note: Survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

The estimated survival rate after 12 months is about 0.2.

To account for clustered data, we need to specify an intraclass correlation in our PSS computations. We can compute a rough estimate of the intraclass correlation assuming a parametric Weibull model as described by Canette (2016). Also see Xie and Waksman (2003) for a nonparametric estimation of the intraclass correlation. For these data, the two methods provide similar values of 0.12 and 0.14, respectively.

We can compute the required number of clusters for a range of intraclass correlation values between 0.04 and 0.2, for example. We specify this range of intraclass correlation values to straddle the rough estimates of 0.12 and 0.14. We specify the control-group survival rate of 0.2 as the command argument,
the hazard ratio of 0.7 in the hratio() option, and because each child has ventilating tubes in both ears, we specify 2 in the m1() and m2() options in the power logrank command below. We also specify the table(rho K1) option to list only the columns that vary in the table.

```
. power logrank 0.2, hratio(0.7) m1(2) m2(2) rho(0.04(0.02)0.2) table(rho k1)
```

Estimated numbers of clusters for two-sample comparison of survivor functions
Cluster randomized design, log-rank test, Freedman method
Ho: HR = 1 versus Ha: HR != 1

<table>
<thead>
<tr>
<th>rho</th>
<th>K1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
<td>89</td>
</tr>
<tr>
<td>0.06</td>
<td>91</td>
</tr>
<tr>
<td>0.08</td>
<td>93</td>
</tr>
<tr>
<td>0.1</td>
<td>94</td>
</tr>
<tr>
<td>0.12</td>
<td>96</td>
</tr>
<tr>
<td>0.14</td>
<td>98</td>
</tr>
<tr>
<td>0.16</td>
<td>100</td>
</tr>
<tr>
<td>0.18</td>
<td>101</td>
</tr>
<tr>
<td>0.2</td>
<td>103</td>
</tr>
</tbody>
</table>

The required number of clusters varies between 89 and 103, with the values of 96 and 98 corresponding to our rough estimates of the intraclass correlation of 0.12 and 0.14.

**Stored results**

`power logrank, cluster` stores the following in `r()`:

**Scalars**
- `r(alpha)`: significance level
- `r(power)`: power
- `r(beta)`: probability of a type II error
- `r(delta)`: effect size
- `r(K1)`: number of clusters in the control group
- `r(K2)`: number of clusters in the experimental group
- `r(kratio)`: ratio of numbers of clusters, K2/K1
- `r(M1)`: cluster size of the control group
- `r(M2)`: cluster size of the experimental group
- `r(mratio)`: ratio of cluster sizes, M2/M1
- `r(N)`: total sample size
- `r(N1)`: sample size of the control group
- `r(N2)`: sample size of the experimental group
- `r(nratio)`: ratio of sample sizes, N2/N1
- `r(nfractional)`: 1 if `nfractional` is specified, 0 otherwise
- `r(onesided)`: 1 for a one-sided test, 0 otherwise
- `r(E)`: total number of events (failures)
- `r(hratio)`: hazard ratio
- `r(lnhratio)`: log hazard-ratio
- `r(s1)`: survival probability in the control group (if specified)
- `r(s2)`: survival probability in the experimental group (if specified)
- `r(Pr_E)`: probability of an event (failure)
- `r(rho)`: intraclass correlation
- `r(CV_cluster)`: coefficient of variation for cluster sizes
- `r(separator)`: number of lines between separator lines in the table
- `r(divider)`: 1 if `divider` is requested in the table, 0 otherwise
- `r(init)`: initial value for estimated parameter
- `r(maxiter)`: maximum number of iterations
Methods and formulas

The computation in a CRD uses the Freedman method based on the asymptotic distribution of the log-rank test statistic. See Methods and formulas in [PSS-2] power logrank for the common notation.

In a CRD, let $K_1$ and $K_2$ be the numbers of clusters in the control and experimental groups, respectively, and $M_1$ and $M_2$ be the cluster sizes of the control and experimental groups, respectively. For unequal cluster sizes, we assume that the cluster sizes are independent and identically distributed and are small relative to the number of clusters. We have $n_1 = K_1M_1$ and $n_2 = K_2M_2$. Let $R$ be the ratio of the numbers of observations, $R_k$ be the ratio of the numbers of clusters, $K_2/K_1$, and $R_m$ be the ratio of the cluster sizes, $M_2/M_1$. Let $n = n_1 + n_2$ be the total number of observations, $K = K_1 + K_2$ be the total number of clusters, $\rho$ be the intraclass correlation, $CV_{cl}$ be the coefficient of variation, and $\bar{M}$ be the average cluster size.

Define $\psi$ as

$$\psi = (R\Delta + 1)/({\Delta - 1})$$

The total number of events required to be observed in a CRD study to ensure a power of $\pi = 1 - \beta$ of the log-rank test to detect the hazard ratio $\Delta$ with significance level $\alpha$, according to Xie and Waksman (2003), is

$$E = \frac{1}{R}(z_{1-\frac{\alpha}{k}} + z_{1-\beta})^2 \left(\frac{R\Delta + 1}{\Delta - 1}\right)^2 \left[1 + \rho\{\bar{M}(1 + CV_{cl}^2) - 1\}\right]$$

where $k = 1$ for the one-sided test and $k = 2$ for the two-sided test.

The required total number of clusters is approximately

$$K = \frac{E}{pE\bar{M}}$$
where $p_E = 1 - (S_1 + RS_2)/(1 + R)$ is the event probability with $S_1$ and $S_2$ defined as the survival probabilities in the control and the experimental groups at the end of the study. Without censoring, $p_E = 1$. Given $R_k$, we can compute $K_1 = K/(1 + R_k)$ and $K_2 = KR_k/(1 + R_k)$.

Given $K_1$ and $K_2$, we can compute the average cluster size $\bar{M}$ as

$$\bar{M} = \frac{1 - \rho}{RKp_E/\{(z_{1-\alpha/k} - z_\beta)\psi\}^2 - \rho(1 + CV_{cl}^2)}$$

We can then compute $M_1 = K\bar{M}/(K_1 + R_mK_2)$ and $M_2 = M_1R_m$.

The power is estimated using the formula

$$\pi = 1 - \beta = \Phi\{|\psi|-1(Rnp_E/[1 + \rho\{(\bar{M}(1 + CV_{cl}^2) - 1]\})^{1/2} - z_{1-\alpha/k}\}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution.

Without censoring, the hazard ratio $\Delta$ is computed as

$$\Delta = \begin{cases} \frac{1 - \frac{R+1}{\sqrt{Rn/\{(z_{1-\alpha/k} - z_\beta)\}^2[1 + \rho(\bar{M}(1 + CV_{cl}^2 - 1)]}} + R}{1 + \frac{R+1}{\sqrt{Rn/\{(z_{1-\alpha/k} - z_\beta)\}^2[1 + \rho(\bar{M}(1 + CV_{cl}^2 - 1)]}} - R} & \text{when } \Delta < 1 \\
\end{cases}$$

with censoring, the hazard ratio $\Delta$ is computed iteratively based on (1) and (2).

References


Also see

[PSS-2] power logrank — Power analysis for the log-rank test

[PSS-2] power — Power and sample-size analysis for hypothesis tests

[PSS-2] power, graph — Graph results from the power command

[PSS-2] power, table — Produce table of results from the power command

[PSS-5] Glossary

[ST] stcox — Cox proportional hazards model

[ST] sts test — Test equality of survivor functions