Title

stpower cox — Sample size, power, and effect size for the Cox proportional hazards model

Syntax

Sample-size determination

> stpower cox [*coef*] [, *options*]

Power determination

> stpower cox [*coef*] , n(*numlist*) [*options*]

Effect-size determination

> stpower cox , n(*numlist*) { power(*numlist*) | beta(*numlist*) } [*options*]

where *coef* is the regression coefficient (effect size) of a covariate of interest, in a Cox proportional hazards model, desired to be detected by a test with a prespecified power. *coef* may be specified either as one number or as a list of values (see [U] **11.1.8 numlist**) enclosed in parentheses.

options	Description
Main	
*alpha(*numlist*)	significance level; default is alpha(0.05)
*power(*numlist*)	power; default is power(0.8)
*beta(*numlist*)	probability of type II error; default is beta(0.2)
*n(*numlist*)	sample size; required to compute power or effect size
*hratio(*numlist*)	hazard ratio (effect size) associated with a one-unit increase in covariate of interest; default is hratio(0.5)
onesided	one-sided test; default is two sided
sd(*#*)	standard deviation of covariate of interest; default is sd(0.5)
r2(*#*)	squared coefficient of multiple correlation with other covariates; default is r2(0)
failprob(*#*)	overall probability of an event (failure) of interest; default is failprob(1), meaning no censoring
wdprob(*#*)	the proportion of subjects anticipated to withdraw from the study; default is wdprob(0)
parallel	treat number lists in starred options as parallel (do not enumerate all possible combinations of values) when multiple values per option are specified

Reporting

<u>hr</u>	report hazard ratio, not coefficient
<u>table</u>	display results in a table with default columns
<u>columns</u>(*colnames*)	display results in a table with specified *colnames* columns
<u>notitle</u>	suppress table title
<u>nolegend</u>	suppress table legend
<u>colw</u>idth(# [# ...])	column widths; default is colwidth(9)
<u>separ</u>ator(#)	draw a horizontal separator line every # lines; default is separator(0) meaning no separator lines
<u>saving</u>(*filename* [, replace])	save the table data to *filename*; use replace to overwrite existing *filename*
<u>nohead</u>er	suppress table header; seldom used
<u>continue</u>	draw a continuation border in the table output; seldom used

[*]Starred options may be specified either as one number or as a list of values (see [U] **11.1.8 numlist**).
noheader and continue are not shown in the dialog box.

colnames	Description
<u>alpha</u>	significance level
<u>power</u>	power
<u>beta</u>	type II error probability
n	total number of subjects
e	total number of events (failures)
hr	hazard ratio
<u>coef</u>	coefficient (log hazard-ratio)
sd	standard deviation
r2	squared multiple-correlation coefficient
pr	overall probability of an event (failure)
w	proportion of withdrawals

By default, the following *colnames* are displayed:
   power, n, e, sd, and alpha are always displayed;
   coef is displayed, unless the hr option is specified, in which case hr is displayed;
   pr if overall probability of an event (failprob()) is specified;
   r2 if squared multiple-correlation coefficient (r2()) is specified; and
   w if withdrawal proportion (wdprob()) is specified.

Menu

Statistics > Survival analysis > Power and sample size

Description

stpower cox estimates required sample size, power, and effect size for survival analyses that use Cox proportional hazards (PH) models. It also reports the number of events (failures) required to be observed in a study. The estimates of sample size or power are obtained for the test of the effect of

one covariate, x_1 (binary or continuous), on time to failure adjusted for other predictors, $x_2, \ldots, x_p$, in a PH model. The command provides options to account for possible correlation between a covariate of interest and other predictors and for withdrawal of subjects from the study. Optionally, the minimal effect size (minimal detectable difference in a regression coefficient, β_1, or hazard ratio) may be obtained for given sample size and power.

You can use `stpower cox` to

- calculate required number of events and sample size when you know power and effect size expressed as a hazard ratio or a coefficient (log hazard-ratio),

- calculate power when you know sample size (number of events) and effect size expressed as a hazard ratio or a coefficient (log hazard-ratio), and

- calculate effect size and display it as a coefficient (log hazard-ratio) or a hazard ratio when you know sample size (number of events) and power.

`stpower cox`'s input parameter, *coef*, is the value β_{1a} of the regression coefficient, β_1, of a covariate of interest, x_1, from a Cox PH model, which is desired to be detected by a test with prespecified power.

Options

      Main

`alpha(`*numlist*`)` sets the significance level of the test. The default is `alpha(0.05)`.

`power(`*numlist*`)` sets the power of the test. The default is `power(0.8)`. If `beta()` is specified, this value is set to be $1-$`beta()`. Only one of `power()` or `beta()` may be specified.

`beta(`*numlist*`)` sets the probability of a type II error of the test. The default is `beta(0.2)`. If `power()` is specified, this value is set to be $1-$`power()`. Only one of `beta()` or `power()` may be specified.

`n(`*numlist*`)` specifies the number of subjects in the study to be used to compute the power of the test or the minimal effect size (minimal detectable value of the regression coefficient, β_1, or hazard ratio) if `power()` or `beta()` is also specified.

`hratio(`*numlist*`)` specifies the hazard ratio associated with a one-unit increase in the covariate of interest, x_1, when other covariates are held constant. The default is `hratio(0.5)`. This value defines the minimal clinically significant effect of a covariate on the response to be detected by a test with a certain power, specified in `power()`, in a Cox PH model. If *coef* is specified, `hratio()` is not allowed and the hazard ratio is instead computed as $\exp(coef)$.

`onesided` indicates a one-sided test. The default is two sided.

`sd(#)` specifies the standard deviation of the covariate of interest, x_1. The default is `sd(0.5)`.

`r2(#)` specifies the squared multiple-correlation coefficient between x_1 and other predictors $x_2, \ldots, x_p$ in a Cox PH model. The default is `r2(0)`, meaning that x_1 is independent of other covariates. This option defines the proportion of variance explained by the regression of x_1 on $x_2, \ldots, x_p$ (see [R] **regress**).

`failprob(#)` specifies the overall probability of a subject failing (or experiencing an event of interest, or not being censored) in the study. The default is `failprob(1)`, meaning that all subjects experience an event (or fail) in the study; that is, no censoring of subjects occurs.

`wdprob(#)` specifies the proportion of subjects anticipated to withdraw from a study. The default is `wdprob(0)`. `wdprob()` may not be combined with `n()`.

parallel reports results sequentially (in parallel) over the list of numbers supplied to options allowing *numlist*. By default, results are computed over all combinations of the number lists in the following order of nesting: alpha(), hratio() or list of coefficients *coef*, power() or beta(), and n(). This option requires that options with multiple values each contain the same number of elements.

⌐ Reporting ⌐

hr specifies that the hazard ratio be displayed rather than the regression coefficient. This option affects how results are displayed and not how they are estimated.

table displays results in a tabular format and is implied if any number list contains more than one element. This option is useful if you are producing results one case at a time and wish to construct your own custom table by using a forvalues loop.

columns(*colnames*) specifies results in a table with specified *colnames* columns. The order of columns in the output table is the same as the order of *colnames* specified in columns(). Column names in columns() must be space-separated.

notitle prevents the table title from displaying.

nolegend prevents the table legend from displaying and column headers from being marked.

colwidth(# [# ...]) specifies column widths. The default is 9 for all columns. The number of specified values may not exceed the number of columns in the table. A missing value (.) may be specified for any column to indicate the default width (9). If fewer widths are specified than the number of columns in the table, the last width specified is used for the remaining columns.

separator(#) specifies how often separator lines should be drawn between rows of the table. The default is separator(0), meaning that no separator lines should be displayed.

saving(*filename* [, replace]) creates a Stata data file (.dta file) containing the table values with variable names corresponding to the displayed *colnames*. replace specifies that *filename* be overwritten if it exists. saving() is appropriate only with tabular output.

The following options are available with stpower cox but are not shown in the dialog box:

noheader prevents the table header from displaying. This option is useful when the command is issued repeatedly, such as within a loop. noheader implies notitle.

continue draws a continuation border at the bottom of the table. This option is useful when the command is issued repeatedly within a loop.

Remarks and examples   stata.com

Remarks are presented under the following headings:

> *Introduction*
> *Computing sample size in the absence of censoring*
> *Computing sample size in the presence of censoring*
> *Link to the sample-size and power computation for the log-rank test*
> *Power and effect-size determination*
> *Performing the analysis with the Cox PH model*

Introduction

Consider a survival study for which the goal is to investigate the effect of a covariate of interest, x_1, on time to failure, possibly adjusted for other predictors, $x_2, \ldots, x_p$, using the Cox proportional hazards model (Cox 1972). For continuous x_1, the effect is measured as a hazard ratio, Δ, or a log hazard-ratio, $\ln(\Delta)$, associated with a one-unit increase in x_1 when the other covariates, $x_2, \ldots, x_p$, are held constant. For a binary predictor, the effect is a ratio of hazards or log hazards corresponding to the two categories of x_1 when other covariates are held constant. In both cases, to measure the effect of a covariate, a test of a hazard or a log hazard-ratio is performed.

In a Cox PH model, the hazard function is assumed to be

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \ldots + \beta_p x_p)$$

where no distributional assumption is made about the baseline hazard, $h_0(t)$. Under this assumption, the regression coefficient, β_1, is the log hazard-ratio, $\ln(\Delta)$, associated with a one-unit increase in x_1 when the other predictors are held constant, and the exponentiated regression coefficient, $\exp(\beta_1)$, is the hazard ratio, Δ. Therefore, the effect of x_1 on time to failure can be investigated by performing an appropriate test based on the partial likelihood (Hosmer, Lemeshow, and May 2008; Klein and Moeschberger 2003) for the regression coefficient, β_1, from a Cox model. Negative values of β_1 correspond to the reduction in hazard for a one-unit increase in x_1, and, conversely, positive values correspond to the increase in hazard for a one-unit increase in x_1.

stpower cox provides the estimates of sample size or power for a test of the regression coefficient, β_1, with the null hypothesis H_0: $(\beta_1, \beta_2, \ldots, \beta_p) = (0, \beta_2, \ldots, \beta_p)$ against the alternative H_a: $(\beta_1, \beta_2, \ldots, \beta_p) = (\beta_{1a}, \beta_2, \ldots, \beta_p)$. The methods used are derived for the score test of H_0 versus H_a. In practice, however, the obtained results may be used in the context of the Wald test as well because the two tests usually lead to the same conclusions about the significance of the regression coefficient. Refer to *The conditional versus unconditional approaches* in [ST] **stpower exponential** for more details about the results based on conditional and unconditional tests. From now on, we will refer to H_a as H_a: $\beta_1 = \beta_{1a}$ for simplicity.

stpower cox implements the method of Hsieh and Lavori (2000) for the sample-size and power computation, which reduces to the method of Schoenfeld (1983) for a binary covariate. The sample size is related to the power of a test through the number of events observed in the study; that is, for a fixed number of events, the power of a test is independent of the sample size. As a result, the sample size is estimated as the number of events divided by the overall probability of a subject failing in the study.

The argument *coef* or hratio() may be used to specify the effect size desired to be detected by a test. If argument *coef* is omitted, the value of the log of the hazard ratio specified in option hratio() or the log of the default hazard-ratio value of 0.5 is used to compute β_{1a}. If argument *coef* is specified, then hratio() is not allowed and the hazard ratio is computed as $\exp(coef)$.

If power determination is desired, then sample size n() must be specified. Otherwise, sample-size determination is assumed with power(0.8) (or, equivalently, beta(0.2)). The default setting for power or, alternatively, the probability of a type II error, a failure to reject the null hypothesis when the alternative hypothesis is true, may be changed by using power() or beta(), respectively. If both n() and power() (or beta()) are specified, then the value of the regression coefficient, β_{1a} (or hazard ratio if the hr option is specified), which can be detected by a test with requested power() for fixed sample size n(), is computed.

The default probability of a type I error, a rejection of the null hypothesis when the null hypothesis is true, of a test is 0.05 but may be changed by using the alpha() option. One-sided tests may be requested by using onesided. By default, no censoring, no correlation between x_1 and other predictors, and no withdrawal of subjects from the study are assumed. This may be changed by specifying failprob(), r2(), and wdprob(), respectively.

Optionally, the results may be displayed in a table by using `table` or `columns()`, as demonstrated in [ST] **stpower**. Refer to [ST] **stpower** and example 7 in *Power and effect-size determination* of [ST] **stpower logrank** to see how to obtain a graph of a power curve.

Computing sample size in the absence of censoring

First, consider a type I study in which all subjects fail by the end of the study (no censoring). Then the required sample size is the same as the number of events required to be observed in a study.

▷ Example 1: Sample size for a model with a binary covariate of interest

Consider a survival study for which the goal is to investigate the effect of a treatment on survival times of subjects. The covariate of interest is binary with levels defining whether a subject receives the treatment (the experimental group) or a placebo (the control or placebo group). Prior to conducting the study, investigators need an estimate of the sample size that ensures that a ratio of hazards of the experimental group to the control group of 0.5 ($\beta_{1a} = \ln(0.5) = -0.6931$) can be detected with a power of 80% with a two-sided, 0.05-level test. Under 1:1 randomization, a subject has a 50% chance of receiving the treatment. The corresponding binary covariate follows a Bernoulli distribution with the probability of a subject receiving a treatment, p, equal to 0.5. As such, the standard deviation of the covariate is $\{p(1-p)\}^{1/2} = 0.5$. Because these study parameters correspond to default values of `stpower cox`, to obtain the sample size for the above study we simply type

```
. stpower cox
Estimated sample size for Cox PH regression
Wald test, log-hazard metric
Ho: [b1, b2, ..., bp] = [0, b2, ..., bp]
Input parameters:
        alpha =    0.0500  (two sided)
           b1 =   -0.6931
           sd =    0.5000
        power =    0.8000
Estimated number of events and sample size:
            E =          66
            N =          66
```

Recall that if argument *coef* is omitted, a default value of $\ln(0.5) = -0.6931$ is assumed. From the output, we see that 66 events (failures) are required to be observed in the study to ensure a power of 80% to detect an alternative $H_a: \beta_1 = -0.6931$ using a two-sided test with a 0.05 significance level. Because we have no censoring, a total of 66 subjects is needed in the study to observe 66 events.

One can also request that the results be displayed in the hazard metric by specifying the `hr` option:

```
. stpower cox, hr
Estimated sample size for Cox PH regression
Wald test, hazard metric
Ho: [b1, b2, ..., bp] = [0, b2, ..., bp]
Input parameters:
        alpha =    0.0500  (two sided)
       hratio =    0.5000
           sd =    0.5000
        power =    0.8000
Estimated number of events and sample size:
            E =          66
            N =          66
```

◁

Suppose now that the covariate of interest, x_1, is continuous. Hsieh and Lavori (2000) extend the formula of Schoenfeld (1983) for the number of events to the case when a covariate is continuous. They also relax the assumption of Schoenfeld (1983) about the independence of x_1 of other covariates and provide an adjustment to the estimate of the number of events for possible correlation.

▷ Example 2: Sample size for a model with a continuous covariate of interest

Consider an example from Hsieh and Lavori (2000) of a study of multiple-myeloma patients treated with alkylating agents (Krall, Uthoff, and Harley 1975). Although in the original study of multiple-myeloma patients, 17 of a total of 65 patients are censored; here we assume that all patients die by the end of the study (a type I study, no censoring). Suppose that the covariate of interest, x_1, is the log of the amount of blood urea nitrogen (BUN) measured in a patient. The sample size for a one-sided, 0.05-level test to detect a coefficient (log hazard-ratio) of 1 for a unit increase in x_1 with a power of 80% is required. The standard deviation of x_1 is 0.3126. To obtain an estimate of the sample size, we supply *coef*, 1, as an argument, the sd(0.3126) option, and the onesided option for a one-sided test.

```
. stpower cox 1, sd(0.3126) onesided

Estimated sample size for Cox PH regression
Wald test, log-hazard metric
Ho: [b1, b2, ..., bp] = [0, b2, ..., bp]

Input parameters:

        alpha =    0.0500  (one sided)
           b1 =    1.0000
           sd =    0.3126
        power =    0.8000

Estimated number of events and sample size:

            E =        64
            N =        64
```

The estimate of the required number of events and the sample size is 64.

◁

Based on the derivation in Schoenfeld (1983) and Hsieh and Lavori (2000), sample-size estimates in the above examples may be used if other covariates are also present in the model as long as these covariates are independent of the covariate of interest. The independence assumption holds for randomized studies, but it is not true for nonrandomized studies, often encountered in practice. Also, in many studies, the main covariate of interest will often be correlated with other covariates. For example, age and gender will often be confounded with the covariate of interest, such as smoking. Below we investigate the effect of the confounding factor on the estimate of the required number of events.

▷ Example 3: Sample size when covariates are not independent

Continuing example 2, the effect of a covariate BUN is desired to be adjusted for eight other covariates in the model. Hsieh and Lavori (2000) report the coefficient of determination of $R^2 = 0.1837$ from regression of the log of BUN, x_1, on the eight other covariates.

```
. stpower cox 1, sd(0.3126) onesided r2(0.1837)

Estimated sample size for Cox PH regression
Wald test, log-hazard metric
Ho: [b1, b2, ..., bp] = [0, b2, ..., bp]

Input parameters:
          alpha =    0.0500  (one sided)
             b1 =    1.0000
             sd =    0.3126
          power =    0.8000
             R2 =    0.1837

Estimated number of events and sample size:
              E =         78
              N =         78
```

The number of events required to be observed in a study and, respectively, the number of subjects increase from 64 to 78 after adjusting for the inflation of the variance of the estimate of β_1 because of the correlation with other covariates. The variance of x_1 decreases by the factor $1 - R^2$, so the estimate of the number of events must also be adjusted by a variance inflation factor $\text{VIF} = 1/(1 - R^2)$.

◁

Computing sample size in the presence of censoring

In the previous section, we assumed that all subjects fail by the end of the study. In practice, the study often terminates after a fixed time, T. As a result, some subjects may not experience an event by the end of the study (a type II study). These subjects are censored. To obtain an estimate of the sample size in the presence of censoring, an estimate of the overall probability of a subject not being censored is required. The investigator may already have such an estimate from previous studies, or this probability may be computed as suggested in the literature (Schoenfeld 1983; Lachin and Foulkes 1986; Barthel et al. 2006; Barthel, Royston, and Babiker 2005; also see [ST] **stpower logrank** and [ST] **stpower exponential**).

▷ Example 4: Sample size in the presence of censoring

Consider the study from example 2. In reality, as mentioned earlier, 17 of a total of 65 patients survived until the end of the study. The overall death rate is estimated as $1 - 17/65 = 0.738$.

```
. stpower cox 1, sd(0.3126) onesided failprob(0.738)

Estimated sample size for Cox PH regression
Wald test, log-hazard metric
Ho: [b1, b2, ..., bp] = [0, b2, ..., bp]

Input parameters:
          alpha =    0.0500  (one sided)
             b1 =    1.0000
             sd =    0.3126
          power =    0.8000
      Pr(event) =    0.7380

Estimated number of events and sample size:
              E =         64
              N =         86
```

In the presence of censoring, the number of subjects required in the study increases from 64 to 86. The number of events remains the same (64) because the only change in the study is the presence of censoring, and censoring is assumed to be independent of failure (event) times.

If we also adjust for the correlation between the log of BUN and other covariates, we obtain the estimate of the sample size to be 106:

```
. stpower cox, hratio(2.7182) sd(0.3126) onesided r2(0.1837) failprob(0.738)

Estimated sample size for Cox PH regression
Wald test, log-hazard metric
Ho: [b1, b2, ..., bp] = [0, b2, ..., bp]

Input parameters:
           alpha =     0.0500  (one sided)
              b1 =     1.0000
              sd =     0.3126
           power =     0.8000
       Pr(event) =     0.7380
              R2 =     0.1837

Estimated number of events and sample size:
               E =         78
               N =        106
```

In the above example, we also demonstrate the alternative syntax. Rather than supplying the coefficient (log hazard-ratio) of 1, we use the `hratio()` option to specify the size of the effect expressed as the hazard ratio $\exp(1) = 2.7182$.

◁

❑ Technical note

Supplying the coefficient (log hazard-ratio) of 1 or -1 (or, respectively, a hazard ratio of $\exp(1) = 2.7182$ or $\exp(-1) = 1/2.7182 = 0.36788$) is irrelevant for sample-size and power determination because it results in the same estimates of sample size and power. However, the sign of the coefficient (or the value of the hazard ratio being larger or smaller than one) is important at the analysis stage because it determines the direction of the effect associated with a one-unit increase of a covariate value.

❑

Often, in practice, subjects may withdraw from a study before it terminates. As a result, the information about the subjects' response is lost. The proportion of subjects anticipated to withdraw from a study may be specified by using `wdprob()`. Refer to [ST] **stpower** and *Withdrawal of subjects from the study* in [ST] **stpower logrank** for a more detailed description and an example.

Link to the sample-size and power computation for the log-rank test

When there are no other covariates in a Cox regression model, the score test of the regression coefficient of a binary covariate is the same as the log-rank test (in the absence of tied observations). Powers of the two tests are the same and therefore so are the formulas for the number of events. The formula for the total number of events for a test of a binary covariate in the context of a PH model in the presence of other covariates is derived in Schoenfeld (1983) under the assumption that the covariate of interest is independent of the other covariates. This formula is also the same as the formula for the number of events when the log-rank test is used to compare survivor functions of two groups without covariates (Schoenfeld 1981). Indeed, using `stpower logrank` for the study described in example 1,

```
. stpower logrank, schoenfeld

Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: S1(t) = S2(t)

Input parameters:
      alpha =    0.0500   (two sided)
 ln(hratio) =   -0.6931
      power =    0.8000
         p1 =    0.5000

Estimated number of events and sample sizes:
          E =         66
          N =         66
         N1 =         33
         N2 =         33
```

yields the same estimates of 66 for both the required number of events and the required sample size.

Schoenfeld (1983) notes that although the formulas for the number of events are the same for the two approaches (based on the log-rank test and based on the score test of a regression coefficient from a Cox regression model adjusting for other covariates), the powers are different. Suppose that the two groups defined by a binary covariate follow the PH model. Then, if covariates are ignored, the ratio of hazards will be nonproportional at every time t and the power of the log-rank test will be smaller than the power of the test based on a Cox PH model.

If a covariate of interest is binary, either stpower cox or stpower logrank with the schoenfeld option can be used to obtain the estimate of the sample size or power regardless of the presence of other covariates. However, if covariates are present, it is important to use the appropriate test that adjusts for other covariates when analyzing the data.

Væth and Skovlund (2004) demonstrate that for a continuous covariate, the sample-size formula for the log-rank test (assuming the equal-group allocation) may be used to obtain the sample size or power with the value of the hazard ratio equal to $\exp(2\beta_{1a}\sigma)$. By typing this expression into the sample-size formula for the log-rank test, one obtains the formula derived in Hsieh and Lavori (2000).

For example, we obtain the same estimate of the total number of events as computed in example 2 using stpower logrank with the schoenfeld option and with the value of the hazard ratio equal to $\exp(2\beta_{1a}\sigma) = \exp(2 \times 1 \times 0.3126) = 1.8686$.

```
. stpower log, hratio(1.8686) onesided schoenfeld

Estimated sample sizes for two-sample comparison of survivor functions
Log-rank test, Schoenfeld method
Ho: S1(t) = S2(t)

Input parameters:
      alpha =    0.0500   (one sided)
 ln(hratio) =    0.6252
      power =    0.8000
         p1 =    0.5000

Estimated number of events and sample sizes:
          E =         64
          N =         64
         N1 =         32
         N2 =         32
```

Power and effect-size determination

Suppose that, for some reason, the number of subjects required to ensure a certain power of a test to detect a specified effect size is not achieved by the end of the recruitment phase of a study. Investigators may want to know by how much the power of a test is decreased for the obtained sample size. If the decrease is significant, then what is the minimal effect size that can be detected with an acceptable level of power for this sample size?

▷ Example 5: Power determination

Consider the data of Krall, Uthoff, and Harley (1975) from the study described in example 2. Suppose that we want to test the effect of the log of BUN on patients' survival times adjusted for eight other covariates. In example 4, the required number of patients is estimated to be 106 to ensure a power of 80% of a 0.05 one-sided test to detect a value of 1 in the regression coefficient. This study, however, had only 65 patients. How does this reduction in sample size affect the power of the test to detect the alternative $H_a: \beta_1 = 1$?

```
. stpower cox 1, sd(0.3126) onesided r2(0.1837) failprob(0.738) n(65)
Estimated power for Cox PH regression
Wald test, log-hazard metric
Ho: [b1, b2, ..., bp] = [0, b2, ..., bp]

Input parameters:
        alpha =     0.0500  (one sided)
           b1 =     1.0000
           sd =     0.3126
            N =         65
    Pr(event) =     0.7380
           R2 =     0.1837

Estimated number of events and power:
            E =         48
        power =     0.6222
```

When the sample size decreases from 106 to 65, power decreases from 80% to 62%.

◁

▷ Example 6: Effect-size determination

Continuing the above example: if a power of 62% is unacceptable to investigators, they may want to find out what is the smallest value of the regression coefficient that can be detected with a preserved power of 80%. To obtain this estimate, we specify both the n() and power() options.

```
. stpower cox, sd(0.3126) onesided r2(0.1837) failprob(0.738) n(65) power(0.8)
Estimated coefficient for Cox PH regression
Wald test, log-hazard metric
Ho: [b1, b2, ..., bp] = [0, b2, ..., bp]

Input parameters:
        alpha =     0.0500  (one sided)
           sd =     0.3126
            N =         65
        power =     0.8000
    Pr(event) =     0.7380
           R2 =     0.1837

Estimated number of events and coefficient:
            E =         48
           b1 =    -1.2711
```

Stata reports the estimate of the regression coefficient of -1.2711. We can disregard the sign because, as mentioned earlier, it is irrelevant in the context of sample-size or power determination. Refer to *Methods and formulas* for details.

With only 65 subjects, the smallest change in log hazards for a one-unit increase in the log of BUN, which can be detected with a preserved 80% power, is roughly 1.27, corresponding to a 27% increase in the log hazard-ratio of 1 desired to be detected originally in example 4.

◁

Performing the analysis with the Cox PH model

After the data are collected, one can use stcox and test to fit the Cox PH model and perform a Wald test, as we demonstrate below.

▷ Example 7: Performing a Wald test

We demonstrate how to perform a Wald test for the regression coefficient of the log of BUN from a Cox model using the data from Krall, Uthoff, and Harley (1975) described in example 2. The dataset myeloma.dta consists of 11 variables, described below.

```
. use http://www.stata-press.com/data/r13/myeloma
(Multiple myeloma patients)

. describe

Contains data from http://www.stata-press.com/data/r13/myeloma.dta
  obs:            65                          Multiple myeloma patients
 vars:            11                          11 Feb 2013 19:26
 size:         1,690
```

variable name	storage type	display format	value label	variable label
time	float	%9.0g		Survival time from diagnosis to nearest month + 1
died	byte	%9.0g		0 – Alive, 1 – Dead
lnbun	float	%9.0g		log BUN at diagnosis
hemo	float	%9.0g		Hemoglobin at diagnosis
platelet	byte	%9.0g	normal	Platelets at diagnosis
age	byte	%9.0g		Age (complete years)
lnwbc	float	%9.0g		Log WBC at diagnosis
fracture	byte	%9.0g	present	Fractures at diagnosis
lnbm	float	%9.0g		log % of plasma cells in bone marrow
protein	byte	%9.0g		Proteinuria at diagnosis
scalcium	byte	%9.0g		Serum calcium (mgm%)

```
Sorted by:
```

Prior to using stcox to fit a Cox model, we need to set up the data by using stset (see [ST] **stset**). The analysis-time variable is time and the failure variable is died.

```
. stset time, failure(died)

     failure event:  died != 0 & died < .
obs. time interval:  (0, time]
 exit on or before:  failure

───────────────────────────────────────────────────────────────────

        65  total observations
         0  exclusions

───────────────────────────────────────────────────────────────────

        65  observations remaining, representing
        48  failures in single-record/single-failure data
    1560.5  total analysis time at risk and under observation
                                       at risk from t =          0
                                  earliest observed entry t =     0
                                   last observed exit t =        92
```

We include all nine covariates in the model and perform a fit by using `stcox`. Then we perform a Wald test of H_0: $\beta_1 = 1$ for the coefficient of `lnbun` using `test`.

```
. stcox lnbun hemo platelet age lnwbc fracture lnbm protein scalcium, nohr
         failure _d:  died
   analysis time _t:  time
Iteration 0:   log likelihood = -154.85799
Iteration 1:   log likelihood = -146.68114
Iteration 2:   log likelihood = -146.29446
Iteration 3:   log likelihood = -146.29404
Refining estimates:
Iteration 0:   log likelihood = -146.29404

Cox regression -- Breslow method for ties

No. of subjects =           65                  Number of obs   =          65
No. of failures =           48
Time at risk    =       1560.5
                                                LR chi2(9)      =       17.13
Log likelihood  =   -146.29404                  Prob > chi2     =      0.0468

──────────────┬──────────────────────────────────────────────────────────────
          _t  │     Coef.    Std. Err.      z    P>|z|     [95% Conf. Interval]
──────────────┼──────────────────────────────────────────────────────────────
       lnbun  │   1.798354   .6483293     2.77   0.006     .5276519    3.069056
        hemo  │  -.1263119   .0718333    -1.76   0.079    -.2671026    .0144789
    platelet  │  -.2505915   .5074656    -0.49   0.621    -1.245206    .7440228
         age  │  -.0127949    .019475    -0.66   0.511    -.0509653    .0253755
       lnwbc  │   .3537259   .7131935     0.50   0.620    -1.044108     1.75156
    fracture  │   .3378767   .4072774     0.83   0.407    -.4603722    1.136126
        lnbm  │   .3589346   .4860298     0.74   0.460    -.5936663    1.311535
     protein  │   .0130672   .0261696     0.50   0.618    -.0382243    .0643587
    scalcium  │   .1259479   .1034015     1.22   0.223    -.0767153    .3286112
──────────────┴──────────────────────────────────────────────────────────────

. test lnbun = 1

 ( 1)  lnbun = 1

           chi2(  1) =    1.52
         Prob > chi2 =    0.2182
```

By default, `stcox` reports estimates of hazard ratios and the two-sided tests of the equality of a coefficient to zero. We use the `nohr` option to request estimates of coefficients. From the output table, a one-sided test of H_0: $\beta_1 = 0$ versus H_a: $\beta_1 > 0$ is rejected at a 0.05 level (one-sided p-value is $0.006/2 = 0.003 < 0.05$). The estimate of the log-hazard difference associated with a one-unit increase of `lnbun` is $\widehat{\beta}_1 = 1.8$. From the `test` output, we cannot reject the hypothesis of H_0: $\beta_1 = 1$.

For these data, the observed effect size (coefficient) of 1.8 is large enough for the sample size of 65 to be sufficient to reject the null hypothesis of no effect of the BUN on the survival of subjects

(H_0: $\beta_1 = 0$). However, if the goal of the study were to ensure that the test detects the effect size corresponding to the coefficient of at least 1 with 80% power, a sample of approximately 106 subjects would have been required.

◁

Stored results

stpower cox stores the following in r():

Scalars

r(N)	total number of subjects
r(E)	total number of events (failures)
r(power)	power of test
r(alpha)	significance level of test
r(hratio)	hazard ratio
r(onesided)	1 if one-sided test, 0 otherwise
r(sd)	standard deviation
r(Pr_E)	probability of an event (failure), if specified
r(r2)	squared multiple correlation, if specified
r(w)	proportion of withdrawals, if specified

Macros

r(metric)	displayed metric (log-hazard or hazard)

Methods and formulas

Let β_1 denote the regression coefficient of the covariate of interest, x_1, from a Cox PH model, possibly in the presence of other covariates, $x_2, \ldots, x_p$; and let Δ denote the hazard ratio associated with a one-unit increase of x_1 when other covariates are held constant. Under the PH model, $\beta_1 = \ln(\Delta)$, where $\ln(\Delta)$ is the change in log hazards associated with a one-unit increase in x_1 when other covariates are held constant.

Define E and N to be the total number of events (failures) and the total number of subjects required in the study; σ to be the standard deviation of x_1; p_{E} to be the overall probability of an event (failure); R^2 to be the proportion of variance explained by the regression of x_1 on $x_2, \ldots, x_p$ (or squared multiple-correlation coefficient); w to be the proportion of subjects withdrawn from a study (lost to follow-up); α to be the significance level; β to be the probability of a type II error; and $z_{(1-\alpha/k)}$ and $z_{(1-\beta)}$ to be the $(1 - \alpha/k)$th and the $(1 - \beta)$th quantiles of the standard normal distribution, with $k = 1$ for the one-sided test and $k = 2$ for the two-sided test.

The total number of events required to be observed in a study to ensure a power of $1 - \beta$ of a test to detect the regression coefficient, β_1, with a significance level α, according to Hsieh and Lavori (2000), is

$$E = \frac{(z_{1-\alpha/k} + z_{1-\beta})^2}{\sigma^2 \beta_1^2 (1 - R^2)}$$

For the case of randomized study and a binary covariate x_1, this formula was derived in Schoenfeld (1983). The formula is an approximation and relies on a set of assumptions such as distinct failure times, all subjects completing the course of the study (no withdrawal), and a local alternative under which $\ln(\Delta)$ is assumed to be of order $O(N^{-1/2})$. The formula is derived for the score test but may be applied to other tests (Wald, for example) that are based on the partial likelihood of a Cox model because all these tests are asymptotically equivalent (Schoenfeld 1983; Hosmer, Lemeshow, and May 2008; Klein and Moeschberger 2003).

The total sample size required to observe the total number of events, E, is given by

$$N = \frac{E}{p_{\mathrm{E}}}$$

The estimate of the sample size is rounded up to the nearest integer.

To account for a proportion of subjects, w, withdrawn from a study, a conservative adjustment to the total sample size suggested in the literature (Freedman 1982; Machin and Campbell 2005) is applied as follows:

$$N_w = \frac{N}{1 - w}$$

Withdrawal is assumed to be independent of administrative censoring and failure (event) times.

Power is estimated using the formula

$$1 - \beta = \Phi\left[|\beta_1|\sigma\{Np_{\mathrm{E}}(1 - R^2)\}^{1/2} - z_{1-\alpha/k}\right]$$

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

The estimate of the regression coefficient for a fixed power, $1 - \beta$, and a sample size, N, is computed as

$$\beta_1^2 = \frac{(z_{1-\alpha/k} + z_{1-\beta})^2}{\sigma^2 Np_{\mathrm{E}}(1 - R^2)}$$

Either of the two values $|\beta_1|$ and $-|\beta_1|$ satisfy the above equation. **stpower cox** reports the negative of the two values, which corresponds to the reduction in a hazard of a failure for a one-unit increase in x_1. Similarly, if the **hr** option is used, the corresponding value of the hazard ratio less than 1 is reported to reflect the reduction in hazard for a one-unit increase in x_1.

References

Barthel, F. M.-S., A. G. Babiker, P. Royston, and M. K. B. Parmar. 2006. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Statistics in Medicine* 25: 2521–2542.

Barthel, F. M.-S., P. Royston, and A. G. Babiker. 2005. A menu-driven facility for complex sample size calculation in randomized controlled trials with a survival or a binary outcome: Update. *Stata Journal* 5: 123–129.

Cox, D. R. 1972. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34: 187–220.

Freedman, L. S. 1982. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* 1: 121–129.

Hosmer, D. W., Jr., S. A. Lemeshow, and S. May. 2008. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. 2nd ed. New York: Wiley.

Hsieh, F. Y., and P. W. Lavori. 2000. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials* 21: 552–560.

Klein, J. P., and M. L. Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. New York: Springer.

Krall, J. M., V. A. Uthoff, and J. B. Harley. 1975. A step-up procedure for selecting variables associated with survival. *Biometrics* 31: 49–57.

Lachin, J. M., and M. A. Foulkes. 1986. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* 42: 507–519.

Machin, D., and M. J. Campbell. 2005. *Design of Studies for Medical Research*. Chichester, UK: Wiley.

Schoenfeld, D. A. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68: 316–319.

———. 1983. Sample-size formula for the proportional-hazards regression model. *Biometrics* 39: 499–503.

Væth, M., and E. Skovlund. 2004. A simple approach to power and sample size calculations in logistic regression and Cox regression models. *Statistics in Medicine* 23: 1781–1792.

Also see [ST] **stpower** for more references.

Also see

[ST] **stpower** — Sample size, power, and effect size for survival analysis

[ST] **stpower exponential** — Sample size and power for the exponential test

[ST] **stpower logrank** — Sample size, power, and effect size for the log-rank test

[ST] **stcox** — Cox proportional hazards model

[ST] **sts test** — Test equality of survivor functions

[ST] **Glossary**

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

[R] **test** — Test linear hypotheses after estimation