Title

stata.com

sts generate — Create variables containing survivor and related functions

Description Quick start Menu Syntax
Functions Options Remarks and examples Methods and formulas
References Also see

Description

`sts generate` creates new variables containing the estimated survivor (failure) function, the Nelson–Aalen cumulative hazard (integrated hazard) function, and related functions. See [ST] **sts** for an introduction to this command.

`sts generate` can be used with single- or multiple-record or single- or multiple-failure st data.

Quick start

Create new variable `surv` containing the Kaplan–Meier survivor function using `stset` data

```
sts generate surv = s
```

Create `sesurv` containing the pointwise standard error for the survivor function

```
sts generate sesurv = se(s)
```

Create `surv2` with separate survivor functions for each level of `v1`

```
sts generate surv2 = s, by(v1)
```

Create `surv3` with survivor function adjusted for `v2` $= 0$

```
sts generate surv3 = s, adjustfor(v2)
```

As above, but create `surv4` with stratification by levels of `svar`

```
sts generate surv3 = s, adjustfor(v2) strata(svar)
```

Create `cumhaz` containing the Nelson–Aalen estimate of the cumulative hazard function

```
sts generate cumhaz = na
```

Menu

Statistics > Survival analysis > Summary statistics, tests, and tables > Create survivor, hazard, and other variables

Syntax

```
sts generate newvar =
    { s | se(s) | h | se(lls) | lb(s) | ub(s) | na | se(na) | lb(na) | ub(na) | n | d }
    [ newvar = {...} ... ] [if] [in] [ , options ]
```

options	Description
Options	
`by(`*varlist*`)`	calculate separately for each group formed by *varlist*
`adjustfor(`*varlist*`)`	adjust the estimates to zero values of *varlist*
`strata(`*varlist*`)`	stratify on different groups of *varlist*
`level(#)`	set confidence level; default is `level(95)`

You must `stset` your data before using `sts generate`; see [ST] **stset**.

Functions

Main

`s` produces the Kaplan–Meier product-limit estimate of the survivor function, $\widehat{S}(t)$, or, if `adjustfor()` is specified, the baseline survivor function from a Cox regression model on the `adjustfor()` variables.

`se(s)` produces the Greenwood, pointwise standard error, $\widehat{\text{se}}\{\widehat{S}(t)\}$. This option may not be used with `adjustfor()`.

`h` produces the estimated hazard component, $\Delta H_j = H(t_j) - H(t_{j-1})$, where t_j is the current failure time and t_{j-1} is the previous one. This is mainly a utility function used to calculate the estimated cumulative hazard, $H(t_j)$, yet you can estimate the hazard via a kernel smooth of the ΔH_j; see [ST] **sts graph**. It is recorded at all the points at which a failure occurs and is computed as d_j/n_j, where d_j is the number of failures occurring at time t_j and n_j is the number at risk at t_j before the occurrence of the failures.

`se(lls)` produces $\widehat{\sigma}(t)$, the standard error of $\ln\{-\ln\widehat{S}(t)\}$. This option may not be used with `adjustfor()`.

`lb(s)` produces the lower bound of the confidence interval for $\widehat{S}(t)$ based on $\ln\{-\ln\widehat{S}(t)\}$: $\widehat{S}(t)^{\exp(-z_{\alpha/2}\widehat{\sigma}(t))}$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. This option may not be used with `adjustfor()`.

`ub(s)` produces the upper bound of the confidence interval for $\widehat{S}(t)$ based on $\ln\{-\ln\widehat{S}(t)\}$: $\widehat{S}(t)^{\exp(z_{\alpha/2}\widehat{\sigma}(t))}$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. This option may not be used with `adjustfor()`.

`na` produces the Nelson–Aalen estimate of the cumulative hazard function. This option may not be used with `adjustfor()`.

`se(na)` produces pointwise standard error for the Nelson–Aalen estimate of the cumulative hazard function, $\widehat{H}(t)$. This option may not be used with `adjustfor()`.

`lb(na)` produces the lower bound of the confidence interval for $\widehat{H}(t)$ based on the log-transformed cumulative hazard function. This option may not be used with `adjustfor()`.

`ub(na)` produces the corresponding upper bound. This option may not be used with `adjustfor()`.

`n` produces n_j, the number at risk just before time t_j. This option may not be used with `adjustfor()`.

`d` produces d_j, the number failing at time t_j. This option may not be used with `adjustfor()`.

Options

Options

`by(`*varlist*`)` performs a separate calculation for each by-group. By-groups are identified by equal values of the variables in *varlist*. `by()` may not be combined with `strata()`.

`adjustfor(`*varlist*`)` adjusts the estimate of the survivor (failure) or hazard function to that for 0 values of *varlist*. This option is available only with functions `s` or `h`. See [ST] **sts graph** for an example of how to adjust for values different from 0.

If you specify `adjustfor()` with `by()`, `sts` fits separate Cox regression models for each group, using the `adjustfor()` variables as covariates. The separately calculated baseline survivor functions are then retrieved.

If you specify `adjustfor()` with `strata()`, `sts` fits a stratified-on-group Cox regression model using the `adjustfor()` variables as covariates. The stratified, baseline survivor function is then retrieved.

`strata(`*varlist*`)` requests estimates of the survivor (failure) or hazard functions stratified on variables in *varlist*. It requires specifying `adjustfor()` and may not be combined with `by()`.

`level(#)` specifies the confidence level, as a percentage, for the `lb(s)`, `ub(s)`, `lb(na)`, and `ub(na)` functions. The default is `level(95)` or as set by `set level`; see [U] **20.8 Specifying the width of confidence intervals**.

Remarks and examples

stata.com

`sts generate` is a seldom-used command that gives you access to the calculations listed by `sts list` and graphed by `sts graph`.

Use of this command is demonstrated in [ST] **sts**.

Methods and formulas

See [ST] **sts**.

References

See [ST] **sts** for references.

Also see

[ST] **sts** — Generate, graph, list, and test the survivor and cumulative hazard functions

[ST] **sts graph** — Graph the survivor, hazard, or cumulative hazard function

[ST] **sts list** — List the survivor or cumulative hazard function

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

[ST] **stset** — Declare data to be survival-time data