strate — Tabulate failure rates and rate ratios

Description

strate tabulates rates by one or more categorical variables declared in *varlist*. You can also save an optional summary dataset, which includes event counts and rate denominators, for further analysis or display. The combination of the commands stsplit and strate implements most of, if not all, the functions of the special-purpose person-years programs in widespread use in epidemiology; see [ST] **stsplit**.

Quick start

Table of failure rates using stset data

    strate

Same as above, but calculate failure rates at each level of categorical variable catvar

    strate catvar

Graph rates against catvar

    strate catvar, graph

Table of SMRs per 1,000 with reference rates stored in variable rvar

    strate catvar, per(1000) smr(rvar)

Menu

Statistics > Survival analysis > Summary statistics, tests, and tables > Tabulate failure rates and rate ratios

Syntax

strate [*varlist*] [*if*] [*in*] [, *options*]

options	Description
[Main]	
per(*#*)	units to be used in reported rates
smr(*varname*)	use *varname* as reference-rate variable to calculate SMRs
cluster(*varname*)	cluster variable to be used by the jackknife
jackknife	report jackknife confidence intervals
missing	include missing values as extra categories
graph	graph rates against exposure category
nowhisker	omit confidence intervals from the graph
level(*#*)	set confidence level; default is level(95)
output(*filename*[, replace])	save summary dataset as *filename*; use replace to overwrite existing *filename*
nolist	suppress listed output
[Plot]	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
cline_options	affect rendition of the plotted points
[CI plot]	
ciopts(*rspike_options*)	affect rendition of the confidence intervals (whiskers)
[Add plots]	
addplot(*plot*)	add other plots to the generated graph
[Y axis, X axis, Titles, Legend, Overall]	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

You must stset your data before using strate; see [ST] **stset**.

fweights, iweights, and pweights may be specified using stset; see [ST] **stset**.

Options

[Main]

per(*#*) specifies the units to be used in reported rates. For example, if the analysis time is in years, specifying per(1000) results in rates per 1,000 person-years.

smr(*varname*) specifies a reference-rate variable. strate then calculates SMRs rather than rates. This option will usually follow stsplit to separate the follow-up records by age bands and possibly calendar periods.

cluster(*varname*) defines a categorical variable that indicates clusters of data to be used by the jack-knife. If the jackknife option is selected and this option is not specified, the cluster variable is taken as the id variable defined in the st data. Specifying cluster() implies jackknife.

jackknife specifies that jackknife confidence intervals be produced. This is the default if weights were specified when the dataset was stset.

missing specifies that missing values of the explanatory variables be treated as extra categories. The default is to exclude such observations.

graph produces a graph of the rate against the numerical code used for the categories of *varname*.

nowhisker omits the confidence intervals from the graph.

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

output(*filename*[, replace]) saves a summary dataset in *filename*. The file contains counts of failures and person-time, rates (or SMRs), confidence limits, and all the categorical variables in the *varlist*. This dataset could be used for further calculations or simply as input to the table command; see [R] **table**.

   replace specifies that *filename* be overwritten if it exists. This option is not shown in the dialog box.

nolist suppresses the output. This is used only when saving results to a file specified by output().

___Plot___

marker_options affect the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] ***marker_options***.

marker_label_options specify if and how the markers are to be labeled; see [G-3] ***marker_label_options***.

cline_options affect whether lines connect the plotted points and the rendition of those lines; see [G-3] ***cline_options***.

___CI plot___

ciopts(*rspike_options*) affects the rendition of the confidence intervals (whiskers); see [G-3] ***rspike_options***.

___Add plots___

addplot(*plot*) provides a way to add other plots to the generated graph; see [G-3] ***addplot_option***.

___Y axis, X axis, Titles, Legend, Overall___

twoway_options are any of the options documented in [G-3] ***twoway_options***, excluding by(). These include options for titling the graph (see [G-3] ***title_options***) and for saving the graph to disk (see [G-3] ***saving_option***).

Remarks and examples

strate tabulates the rate, formed from the number of failures divided by the person-time, by different levels of one or more categorical explanatory variables specified by *varlist*. Confidence intervals for the rate are also given. By default, the confidence intervals are calculated using the quadratic approximation to the Poisson log likelihood for the log-rate parameter. However, whenever the Poisson assumption is questionable, jackknife confidence intervals can also be calculated. The jackknife option also allows for multiple records for the same cluster (usually subject).

strate can also calculate and report SMRs if the data have been merged with a suitable file of reference rates.

The summary dataset can be saved to a file specified with the output() option for further analysis or more elaborate graphical display.

If weights were specified when the dataset was stset, strate calculates jackknife confidence intervals by default.

▷ Example 1: Tabulation of failure rates

Using the diet data (Clayton and Hills 1993) described in example 1 of [ST] **stsplit**, we will use strate to tabulate age-specific coronary heart disease (CHD). In this dataset, CHD has been coded as fail = 1, 3, or 13.

We first stset the data: failure codes for CHD are specified; origin is set to date of birth, making age the analysis time; and the scale is set to 365.25, so analysis time is measured in years.

```
. use https://www.stata-press.com/data/r19/diet
(Diet data with dates)
. stset dox, origin(time dob) enter(time doe) id(id) scale(365.25)
> fail(fail==1 3 13)

Survival-time data settings

           ID variable: id
        Failure event: fail==1 3 13
Observed time interval: (dox[_n-1], dox]
    Enter on or after: time doe
    Exit on or before: failure
    Time for analysis: (time-origin)/365.25
               Origin: time dob

─────────────────────────────────────────────────────────────────
        337  total observations
          0  exclusions
─────────────────────────────────────────────────────────────────
        337  observations remaining, representing
        337  subjects
         46  failures in single-failure-per-subject data
  4,603.669  total analysis time at risk and under observation
                                       At risk from t =          0
                             Earliest observed entry t =   30.07529
                                 Last observed exit t =   69.99863
```

Now we stsplit the data into 10-year age bands.

```
. stsplit ageband, at(40(10)70) after(time=dob) trim
(26 observations trimmed because of lower bound)
(418 observations (episodes) created)
```

stsplit added 418 observations to the dataset in memory and generated a new variable, ageband, that identifies each observation's age group.

The CHD rate per 1,000 person-years can now be tabulated for categories of ageband:

```
. strate ageband, per(1000) graph
          Failure _d: fail==1 3 13
    Analysis time _t: (dox-origin)/365.25
              Origin: time dob
  Enter on or after: time doe
         ID variable: id
                Note: ageband<=40 trimmed
Estimated failure rates
Number of records = 729
```

ageband	D	Y	Rate	Lower	Upper
40	6	0.9070	6.6152	2.9719	14.7246
50	18	2.1070	8.5428	5.3823	13.5591
60	22	1.4933	14.7325	9.7007	22.3746

```
Notes: Rate = D/Y = failures/person-time (per 1000).
       Lower and Upper are bounds of 95% confidence intervals.
```
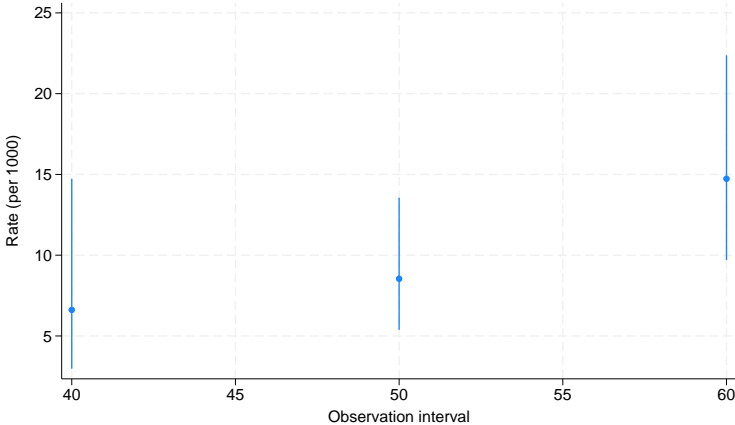

Because we specified the graph option, strate also generated a plot of the estimated rates and confidence intervals.

◁

The SMR for a cohort is the ratio of the total number of observed deaths to the number expected from age-specific reference rates. This expected number can be found by first expanding on age, using stsplit, and then multiplying the person-years in each age band by the reference rate for that band. merge (see [D] **merge**) can be used to add the reference rates to the dataset. Using the smr option to define the variable containing the reference rates, strate calculates SMRs and confidence intervals. You must specify the per() option. For example, if the reference rates were per 100,000 person-years, you would specify per(100000). When reference rates are available by age and calendar period, you must call stsplit twice to expand on both time scales before merging the data with the reference-rate file.

▷ Example 2: Tabulation of SMRs

In smrchd.dta, we have age-specific CHD rates per 1,000 person-years for a reference population. We can merge these data with our current data and use strate to obtain SMRs and confidence intervals.

```
. sort ageband
. merge m:1 ageband using https://www.stata-press.com/data/r19/smrchd
(variable ageband was byte, now float to accommodate using data's values)
    Result                        Number of obs
    ─────────────────────────────────────────
    Not matched                             26
        from master                         26  (_merge==1)
        from using                           0  (_merge==2)
    Matched                                729  (_merge==3)
    ─────────────────────────────────────────

. strate ageband, per(1000) smr(rate)
          Failure _d: fail==1 3 13
    Analysis time _t: (dox-origin)/365.25
              Origin: time dob
   Enter on or after: time doe
          ID variable: id
                 Note: ageband<=40 trimmed
Estimated standardized mortality rates
Reference-rate variable: rate

Number of records = 729

    ┌─────────────────────────────────────────────────────┐
    │ ageband   D       E      SMR    Lower    Upper        │
    │─────────────────────────────────────────────────────│
    │      40   6     5.62   1.0670   0.4793   2.3749        │
    │      50  18    18.75   0.9599   0.6048   1.5235        │
    │      60  22    22.85   0.9629   0.6340   1.4624        │
    └─────────────────────────────────────────────────────┘

 Notes: SMR = D/E = failures/expected failures (per 1000).
        Lower and Upper are bounds of 95% confidence intervals.
```

◁

Acknowledgments

Reference

Clayton, D. G., and M. Hills. 1993. *Statistical Models in Epidemiology*. Oxford: Oxford University Press.

Also see

[ST] **stci** — Confidence intervals for means and percentiles of survival time

[ST] **stir** — Report incidence-rate comparison

[ST] **stmc** — Calculate rate ratios with the Mantel–Cox method

[ST] **stmh** — Calculate rate ratios with the Mantel–Haenszel method

[ST] **stptime** — Calculate person-time, incidence rates, and SMR

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

For suggested citations, see the FAQ on citing Stata documentation.