Title

stata.com

stptime — Calculate person-time, incidence rates, and SMR

Description

`stptime` calculates person-time and incidence rates. `stptime` computes standardized mortality/morbidity ratios (SMRs) after merging the data with a suitable file of standard rates specified with the `using()` option.

Quick start

Person-time and incidence rate using `stset` data

```
stptime
```

Same as above, but tabulate in ten-year intervals from 20 to 50

```
stptime, at(20(10)50)
```

Same as above, but exclude observations less than or equal to 20 or greater than 50

```
stptime, at(20(10)50) trim
```

Same as above, but report rate per 1,000 person-years with two decimal places

```
stptime, at(20(10)50) trim per(1000) dd(2)
```

Person-time and incidence rates for each level of `v1`

```
stptime, by(v1)
```

Standardized mortality ratios in 10-year intervals from 20 to 50 from reference rates `rvar` for lower end-points `lower`, defining each cohort saved in `mydata.dta`

```
stptime, at(20(10)50) smr(lower rvar) using(mydata)
```

Menu

Statistics > Survival analysis > Summary statistics, tests, and tables > Person-time, incidence rates, and SMR

Syntax

```
stptime [if] [, options]
```

options	Description
Main	
`at(`*numlist*`)`	compute person-time at specified intervals; default is to compute overall person-time and incidence rates
`trim`	exclude observations $\leq$ minimum or $>$ maximum of `at()`
`by(`*varname*`)`	compute incidence rates or SMRs by *varname*
Options	
`per(#)`	units to be used in reported rates
`dd(#)`	number of decimal digits to be displayed
`smr(`*groupvar ratevar*`)`	use *groupvar* and *ratevar* in `using()` dataset to calculate SMRs
`using(`*filename*`)`	specify filename to merge that contains `smr()` variables
`level(#)`	set confidence level; default is `level(95)`
`noshow`	do not show st setting information
Advanced	
`jackknife`	jackknife confidence intervals
`title(`*string*`)`	label output table with *string*
`output(`*filename*`[ , replace ])`	save summary dataset as *filename*; use `replace` to overwrite existing *filename*

You must `stset` your data before using `stptime`; see [ST] **stset**.
`by` and `collect` are allowed; see [U] **11.1.10 Prefix commands**.
`fweights`, `iweights`, and `pweights` may be specified using `stset`; see [ST] **stset**.

Options

Main

`at(`*numlist*`)` specifies intervals at which person-time is to be computed. The intervals are specified in analysis time t units. If `at()` is not specified, overall person-time and incidence rates are computed.

If, for example, you specify `at(5(5)20)` and the `trim` option is not specified, person-time is reported for the intervals $t = (0 - 5]$, $t = (5 - 10]$, $t = (10 - 15]$, and $t = (15 - 20]$.

`trim` specifies that observations less than or equal to the minimum or greater than the maximum value listed in `at()` be excluded from the computations.

`by(`*varname*`)` specifies a categorical variable by which incidence rates or SMRs are to be computed.

Options

`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.

`dd(#)` specifies the maximum number of decimal digits to be reported for rates, ratios, and confidence intervals. This option affects only how values are displayed, not how they are calculated.

`smr(`*groupvar ratevar*`)` specifies two variables in the `using()` dataset. The *groupvar* identifies the age-group or calendar-period variable used to match the data in memory and the `using()` dataset. The *ratevar* variable contains the appropriate reference rates. `stptime` then calculates SMRs rather than incidence rates.

`using(`*filename*`)` specifies the filename that contains a file of standard rates that is to be merged with the data so that SMRs can be calculated.

`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**.

`noshow` prevents `stptime` from showing the key st variables. This option is seldom used because most people type `stset, show` or `stset, noshow` to set whether they want to see these variables mentioned at the top of the output of every st command; see [ST] **stset**.

Advanced

`jackknife` specifies that jackknife confidence intervals be produced. This is the default if `pweights` or `iweights` were specified when the dataset was `stset`.

`title(`*string*`)` replaces the default "Person-time" label on the output table with *string*.

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

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

Remarks and examples

stata.com

`stptime` computes and tabulates the person-time and incidence rate (formed from the number of failures divided by the person-time). If you use the `by()` option, this will be calculated by different levels of one or more categorical explanatory variables specified by *varname*. 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, such as when `pweights` or `iweights` are used, jackknife confidence intervals can also be calculated.

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

If `pweights` or `iweights` were specified when the dataset was `stset`, `stptime` calculates jackknife confidence intervals by default.

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

Example 1

We begin with a simple fictitious example from Clayton and Hills (1993, 42). Thirty subjects were monitored until the development of a particular disease. Here are the data for the first five subjects:

```
. use https://www.stata-press.com/data/r18/stptime
. list in 1/5
```

	id	year	fail
1.	1	19.6	1
2.	2	10.8	1
3.	3	14.1	1
4.	4	3.5	1
5.	5	4.8	1

The `id` variable identifies the subject, `year` records the time to failure in years, and `fail` is the failure indicator, which is 1 for all 30 subjects in the data. To use `stptime`, we must first `stset` the data.

```
. stset year, fail(fail) id(id)
Survival-time data settings
           ID variable: id
         Failure event: fail!=0 & fail<.
Observed time interval: (year[_n-1], year]
     Exit on or before: failure
--------------------------------------------------------------------------------
         30  total observations
          0  exclusions
--------------------------------------------------------------------------------
         30  observations remaining, representing
         30  subjects
         30  failures in single-failure-per-subject data
      261.9  total analysis time at risk and under observation
                                                At risk from t =         0
                                     Earliest observed entry t =         0
                                          Last observed exit t =      36.5
```

We can use `stptime` to obtain the overall person-time of observation and disease incidence rate.

```
. stptime, title(Person-years)
        Failure _d: fail
  Analysis time _t: year
       ID variable: id
Estimated person-time and incidence rate
```

Cohort	Person-years	Failures	Rate	[95% conf.	interval]
Total	261.9	30	.11454754	.08009	.1638299

The total 261.9 person-years reported by `stptime` matches what `stset` reported as total analysis time at risk. `stptime` computed an incidence rate of 0.11454754 per person-year. In epidemiology, incidence rates are often presented per 1,000 person-years. We can do this by specifying `per(1000)`.

```
. stptime, title(Person-years) per(1000)
        Failure _d: fail
  Analysis time _t: year
       ID variable: id

Estimated person-time and incidence rate
```

Cohort	Person-years	Failures	Rate	[95% conf.	interval]
Total	261.9	30	114.54754	80.09001	163.8299

More interesting would be to compare incidence rates at 10-year intervals. We will specify `dd(4)` to display rates to four decimal places.

```
. stptime, per(1000) at(0(10)40) dd(4)
        Failure _d: fail
  Analysis time _t: year
       ID variable: id

Estimated person-time and incidence rates
```

Cohort	Person-time	Failures	Rate	[95% conf.	interval]
(0 - 10]	188.8000	18	95.3390	60.0676	151.3215
(10 - 20]	55.1000	10	181.4882	97.6506	337.3044
(20 - 30]	11.5000	1	86.9565	12.2490	617.3106
> 30	6.5000	1	153.8462	21.6713	1092.1648
Total	261.9000	30	114.5475	80.0900	163.8299

◁

▷ Example 2

Using the diet data (Clayton and Hills 1993) described in example 1 of [ST] **stsplit**, we will use `stptime` to tabulate age-specific person-years and coronary heart disease (CHD) incidence rates. 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/r18/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
```

The incidence of CHD per 1,000 person-years can be tabulated in 10-year intervals.

```
. stptime, per(1000) at(40(10)70) trim
        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: _group<=40 trimmed
Estimated person-time and incidence rates
    Cohort |  Person-time   Failures        Rate   [95% conf. interval]
-----------+-------------------------------------------------------------
(40  -  50]    907.00616          6   6.6151701   2.971936    14.72457
(50  -  60]    2107.0418         18   8.5427828   5.382317    13.55906
(60  -  70]    1493.2923         22   14.732548   9.700656    22.37457
-----------+-------------------------------------------------------------
     Total    4507.3402         46   10.205575   7.644246    13.62512
```

◁

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 multiplying the person-time in each cohort by the reference rate for that cohort. Using the `smr` option to define the cohort variable and reference rate variable in the `using()` dataset, `stptime` calculates SMRs and confidence intervals. You must specify the `per()` option. For example, if the reference rates were per 100,000, you would specify `per(100000)`.

▷ Example 3

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 stptime to obtain SMRs and confidence intervals.

```
. stptime, smr(ageband rate) using(https://www.stata-press.com/data/r18/smrchd)
> per(1000) at(40(10)70) trim
        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: _group<=40 trimmed
Estimated person-time and standardized mortality ratios
```

Cohort	Person-time	Observed failures	Expected failures	SMR	[95% conf.	interval]
(40 - 50]	907.00616	6	5.62344	1.067	.4793445	2.374931
(50 - 60]	2107.0418	18	18.7527	.95986	.6047547	1.52349
(60 - 70]	1493.2923	22	22.8474	.96291	.6340298	1.46239
Total	4507.3402	46	47.2235	.97409	.7296205	1.300477

The stptime command can also calculate person-time and incidence rates or SMRs by categories of the explanatory variable. In our diet data, the variable hienergy is coded 1 if the total energy consumption is more than 2.75 Mcal and 0 otherwise. We want to compute the person-years and incidence rates for these two levels of hienergy.

```
. stptime, by(hienergy) per(1000)
        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
Estimated person-time and incidence rates
```

hienergy	Person-time	Failures	Rate	[95% conf.	interval]
0	2059.4305	28	13.595992	9.387478	19.69123
1	2544.2382	18	7.0748093	4.457431	11.2291
Total	4603.6687	46	9.9920309	7.484296	13.34002

We can also compute the incidence rate for the two levels of hienergy and the three previously defined age cohorts:

```
. stptime, by(hienergy) per(1000) at(40(10)70) trim
        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
```

Estimated person-time and incidence rates

hienergy	Person-time	Failures	Rate	[95% conf.	interval]
0					
(40 - 50]	346.87474	2	5.76577	1.442006	23.05407
(50 - 60]	979.34018	12	12.253148	6.958681	21.57587
> 60	699.13758	14	20.024671	11.85966	33.81104
1					
(40 - 50]	560.13142	4	7.1411813	2.680213	19.02702
(50 - 60]	1127.7016	6	5.3205566	2.390317	11.84292
> 60	794.15469	8	10.073604	5.037786	20.14327
Total	4507.3402	46	10.205575	7.644246	13.62512

Or we can compute the corresponding SMR:

```
. stptime, smr(ageband rate) using(https://www.stata-press.com/data/r18/smrchd)
> by(hienergy) per(1000) at(40(10)70) trim
        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
```

Estimated person-time and standardized mortality ratios

hienergy	Person-time	Observed failures	Expected failures	SMR	[95% conf.	interval]
0						
(40 - 50]	346.87474	2	2.15062	.9299629	.2325815	3.718399
(50 - 60]	979.34018	12	8.71613	1.376758	.7818743	2.424256
> 60	699.13758	14	10.6968	1.308802	.7751411	2.209872
1						
(40 - 50]	560.13142	4	3.47281	1.151803	.4322924	3.068875
(50 - 60]	1127.7016	6	10.0365	.5978154	.2685749	1.330665
> 60	794.15469	8	12.1506	.6584055	.329267	1.316554
Total	4507.3402	46	47.2235	.9740917	.7296205	1.300477

◁

Video example

How to calculate incidence rates and incidence-rate ratios

Stored results

stptime stores the following in r():

Scalars

r(ptime)	person-time
r(failures)	observed failures
r(rate)	failure rate
r(expected)	expected number of failures
r(smr)	standardized mortality ratio
r(lb)	lower bound for SMR
r(ub)	upper bound for SMR

References

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

Rutherford, M. J., P. C. Lambert, and J. Thompson. 2010. Age–period–cohort modeling. *Stata Journal* 10: 606–627.

Also see

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

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

[ST] **strate** — Tabulate failure rates and rate ratios

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

[ST] **stsplit** — Split and join time-span records

[R] **Epitab** — Tables for epidemiologists