stdescribe — Describe survival-time data

Description

`stdescribe` reports the characteristics of a survival-time dataset. The report includes the number of subjects and per-subject summary statistics related to the number of records, entry and exit times, gaps in the data, time at risk, and number of failures.

`stdescribe` can be used with single- or multiple-record and single- or multiple-failure st data.

Quick start

Report characteristics of a survival-time dataset using `stset` data

```
stdescribe
```

Describe only data with `v1` = 1

```
stdescribe if v1==1
```

Compute weighted statistics using the weight specified in `stset`

```
stdescribe, weight
```

Menu

Statistics > Survival analysis > Summary statistics, tests, and tables > Describe survival-time data

Syntax

`stdescribe` [*if*] [*in*] [, `weight` `noshow`]

You must `stset` your data before using `stdescribe`; 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

`weight` specifies that the summary use weighted rather than unweighted statistics. `weight` does nothing unless you specified a weight when you `stset` the data. The `weight` option and the ability to ignore weights are unique to `stdescribe`. The purpose of `stdescribe` is to describe the data in a computer sense—the number of records, etc.—and for that purpose, the weights are best ignored.

`noshow` prevents `stdescribe` 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**.

Remarks and examples

Here is an example of stdescribe with single-record survival data:

```
. use https://www.stata-press.com/data/r19/page2
. stdescribe
        Failure _d: dead
  Analysis time _t: time
```

Category	Total	Per subject Mean	Min	Median	Max
Number of subjects	40				
Number of records	40	1	1	1	1
Entry time (first)		0	0	0	0
Exit time (final)		227.95	142	231	344
Subjects with gap	0				
Time on gap	0				
Time at risk	9118	227.95	142	231	344
Failures	36	.9	0	1	1

There is one record per subject. The purpose of this summary is not analysis—it is to describe how the data are arranged. We can quickly see that there is one record per subject (the number of subjects equals the number of records, but if there is any doubt, the minimum and maximum number of records per subject is 1), that all the subjects entered at time 0, that the subjects exited between times 142 and 344 (median 231), that there are no gaps (as there could not be if there is only one record per subject), that the total time at risk is 9,118 (distributed reasonably evenly across the subjects), and that the total number of failures is 36 (with a maximum of 1 failure per subject).

Here is a description of the multiple-record Stanford heart transplant data that we introduced in [ST] **stset**:

```
. use https://www.stata-press.com/data/r19/stan3
(Heart transplant data)
. stdescribe
        Failure _d: died
  Analysis time _t: t1
       ID variable: id
```

Category	Total	Per subject Mean	Min	Median	Max
Number of subjects	103				
Number of records	172	1.669903	1	2	2
Entry time (first)		0	0	0	0
Exit time (final)		310.0786	1	90	1799
Subjects with gap	0				
Time on gap	0	.	.	.	.
Time at risk	31938.1	310.0786	1	90	1799
Failures	75	.7281553	0	1	1

Here patients have one or two records. Although this is not revealed by the output, a patient has one record if the patient never received a heart transplant and two if the patient did receive a transplant; the first reflects the patient's survival up to the time of transplantation and the second their subsequent survival:

```
. stset, noshow              /* to not show the st marker variables */
. stdescribe if !transplant
```

		Per subject			
Category	Total	Mean	Min	Median	Max
Number of subjects	34				
Number of records	34	1	1	1	1
Entry time (first)		0	0	0	0
Exit time (final)		96.61765	1	21	1400
Subjects with gap	0				
Time on gap	0	.	.	.	.
Time at risk	3285	96.61765	1	21	1400
Failures	30	.8823529	0	1	1

```
. stdescribe if transplant
```

		Per subject			
Category	Total	Mean	Min	Median	Max
Number of subjects	69				
Number of records	138	2	2	2	2
Entry time (first)		0	0	0	0
Exit time (final)		415.2623	5.1	207	1799
Subjects with gap	0				
Time on gap	0	.	.	.	.
Time at risk	28653.1	415.2623	5.1	207	1799
Failures	45	.6521739	0	1	1

Finally, here are the results of `stdescribe` from multiple-failure data:

```
. use https://www.stata-press.com/data/r19/mfail2
. stdescribe
```

		Per subject			
Category	Total	Mean	Min	Median	Max
Number of subjects	926				
Number of records	1734	1.87257	1	2	4
Entry time (first)		0	0	0	0
Exit time (final)		470.6857	1	477	960
Subjects with gap	6				
Time on gap	411	68.5	16	57.5	133
Time at risk	435444	470.2419	1	477	960
Failures	808	.8725702	0	1	3

The maximum number of failures per subject observed is three, although 50% had just one failure, and six subjects have gaps in their histories.

Video example

How to describe and summarize survival data

Stored results

stdescribe stores the following in r():

Scalars

r(N_sub)	number of subjects	r(gap)	total gap, if gap
r(N_total)	number of records	r(gap_min)	minimum gap, if gap
r(N_min)	minimum number of records	r(gap_mean)	mean gap, if gap
r(N_mean)	mean number of records	r(gap_med)	median gap, if gap
r(N_med)	median number of records	r(gap_max)	maximum gap, if gap
r(N_max)	maximum number of records	r(tr)	total time at risk
r(t0_min)	minimum first entry time	r(tr_min)	minimum time at risk
r(t0_mean)	mean first entry time	r(tr_mean)	mean time at risk
r(t0_med)	median first entry time	r(tr_med)	median time at risk
r(t0_max)	maximum first entry time	r(tr_max)	maximum time at risk
r(t1_min)	minimum final exit time	r(N_fail)	number of failures
r(t1_mean)	mean final exit time	r(f_min)	minimum number of failures
r(t1_med)	median final exit time	r(f_mean)	mean number of failures
r(t1_max)	maximum final exit time	r(f_med)	median number of failures
r(N_gap)	number of subjects with gap	r(f_max)	maximum number of failures

Reference

Cleves, M. A., W. W. Gould, and Y. V. Marchenko. 2016. *An Introduction to Survival Analysis Using Stata*. Rev. 3rd ed. College Station, TX: Stata Press.

Also see

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

[ST] **stsum** — Summarize survival-time data

[ST] **stvary** — Report variables that vary over time