snapspan — Convert snapshot data to time-span data

Syntax

```
snapspan idvar timevar varlist [, generate(newt0var) replace]
```

Menu

Statistics > Survival analysis > Setup and utilities > Convert snapshot data to time-span data

Description

snapspan converts snapshot data to time-span data. See Snapshot and time-span datasets below for a description of snapshot and time-span data. Time-span data are required for use with survival analysis commands, such as stcox, streg, and stset.

- `idvar` records the subject ID and may be string or numeric.
- `timevar` records the time of the snapshot; it must be numeric and may be recorded on any scale: date, hour, minute, second, etc.
- `varlist` are the “event” variables, meaning that they occur at the instant of `timevar`. `varlist` can also include retrospective variables that are to apply to the time span ending at the time of the current snapshot. The other variables are assumed to be measured at the time of the snapshot and thus apply from the time of the snapshot forward. See Specifying varlist below.

Options

- `generate(newt0var)` adds `newt0var` to the dataset containing the entry time for each converted time-span record.
- `replace` specifies that it is okay to change the data in memory, even though the dataset has not been saved on disk in its current form.

Remarks and examples

Remarks are presented under the following headings:

- Snapshot and time-span datasets
- Specifying varlist

Snapshot and time-span datasets

snapspan converts a snapshot dataset to a time-span dataset. A snapshot dataset records a subject `id`, a `time`, and then other variables measured at the `time`:
Snapshot datasets:

idvar  timevar  x1  x2  ...
47     12        5   27  ...
47     42        5   18  ...
47     55        5   19  ...

idvar  datevar  x1  x2  ...
122    14jul1998 5   27  ...
122    12aug1998 5   18  ...
122    08sep1998 5   19  ...

idvar  year  x1  x2  ...
122    1994  5   27  ...
122    1995  5   18  ...
122    1997  5   19  ...

A time-span dataset records a span of time \((time0, time1]\):

\[
\begin{array}{c}
\text{some variables assumed} \\
\text{to occur at time1} \\
\leftarrow \text{other variables assumed constant over span} \\
\rightarrow \text{time}
\end{array}
\]

\(time0\)  \(time1\)

Time-span data are required, for instance, by \texttt{stset} and the \texttt{st} system. The variables assumed to occur at \texttt{time1} are the failure or event variables. All the other variables are assumed to be constant over the span.

Time-span datasets:

idvar  time0  time1  x1  x2  ...  event
47     0      12    5   13  ...   0
47     12     42    5   27  ...   0
47     42     55    5   18  ...   1

idvar  time0  time1  x1  x2  ...  event
122    01jan1998 14jul1998 5  13  ...   0
122    14jul1998 12aug1998 5  27  ...   0
122    12aug1998 08sep1998 5  18  ...   1

idvar  time0  time1  x1  x2  ...  event
122    1993    1994  5   13  ...   0
122    1994    1995  5   27  ...   0
122    1995    1997  5   18  ...   1

To convert snapshot data to time-span data, you need to distinguish between event and nonevent variables. Event variables happen at an instant.

Say that you have a snapshot dataset containing variable \(e\) recording an event (\(e = 1\) might record surgery, death, becoming unemployed, etc.) and the rest of the variables—call them \(x1\), \(x2\), etc.—recording characteristics (such as sex, birth date, blood pressure, or weekly wage). The same data, in snapshot and time-span form, would be
In snapshot form:

<table>
<thead>
<tr>
<th>id</th>
<th>time</th>
<th>x1</th>
<th>x2</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>a1</td>
<td>b1</td>
<td>e1</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>a2</td>
<td>b2</td>
<td>e2</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>a3</td>
<td>b3</td>
<td>e3</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>a4</td>
<td>b4</td>
<td>e4</td>
</tr>
</tbody>
</table>

In time-span form:

<table>
<thead>
<tr>
<th>id</th>
<th>time0</th>
<th>time</th>
<th>x1</th>
<th>x2</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.</td>
<td>5</td>
<td>.</td>
<td>.</td>
<td>e1</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>7</td>
<td>a1</td>
<td>b1</td>
<td>e2</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>9</td>
<td>a2</td>
<td>b2</td>
<td>e3</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>11</td>
<td>a3</td>
<td>b3</td>
<td>e4</td>
</tr>
</tbody>
</table>

snapspan converts data from the form on the left to the form on the right:

```
    . snapspan id time e, generate(time0) replace
```

The form on the right is suitable for use by `stcox` and `stset` and the other survival analysis commands.

Specifying varlist

The `varlist`—the third variable on—specifies the “event” variables.

In fact, the `varlist` specifies the variables that apply to the time span ending at the time of the current snapshot. The other variables are assumed to be measured at the time of the snapshot and thus apply from the time of the snapshot forward.

Thus `varlist` should include retrospective variables.

For instance, say that the snapshot recorded `bp`, blood pressure; `smokes`, whether the patient smoked in the last 2 weeks; and `event`, a variable recording examination, surgery, etc. Then `varlist` should include `smokes` and `event`. The remaining variables, `bp` and the rest, would be assumed to apply from the time of the snapshot forward.

Suppose that the snapshot recorded `ecs`, employment change status (hired, fired, promoted, etc.); `wage`, the current hourly wage; and `ms`, current marital status. Then `varlist` should include `esc` and `ms` (assuming snapshot records are not generated for reason of `ms` change). The remaining variables, `wage` and the rest, would be assumed to apply from the time of the snapshot forward.

Also see

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