Description

In this example, we demonstrate how to fit a survival model to data that are both left-truncated and right-censored.

```
. use https://www.stata-press.com/data/r16/gsem_diet
(Diet data with dates)
. describe
Contains data from https://www.stata-press.com/data/r16/gsem_diet.dta
obs: 337 Diet data with dates
vars: 11 16 Jan 2019 11:24
(_dta has notes)

storage  display value             variable name  type format label
variable name type format label
id       int %9.0g Subject identity number
fail     byte %8.0g Outcome (CHD = 1 3 13)
job      byte %8.0g Occupation
month    byte %8.0g month of survey
energy   float %9.0g Total energy (1000kcals/day)
height   float %9.0g Height (cm)
weight   float %9.0g Weight (kg)
hie       byte %9.0g Indicator for high energy
doee     int %td Date of entry
doex     int %td Date of exit
dobe     int %td Date of birth
```

Sorted by: id

. notes
_dta:

See *Structural models 7: Survival models* in [SEM] *Intro 5* for background.

Remarks and examples

Remarks are presented under the following headings:

*Censoring and truncation*
*Using stset to declare survival characteristics*
*Fitting the loglogistic model*
*Fitting the model with the Builder*
Censoring and truncation

Survival datasets often include observations that are right-censored or left-truncated or both. When fitting survival models with family(exponential), family(gamma), family(loglogistic), family(lognormal), and family(weibull), we can include the suboption ltruncated(), which specifies a left-truncation point, and the suboption failure(), which indicates whether an observation records a failure or whether it was censored. For instance,

```
gsem failtime ..., failure(weibull, failure(failed) ltruncated(enter))
```

fits a Weibull model for time to failure (failtime) where a failure is observed for observations with failed = 1 and observations with failed = 0 are censored. In addition, observations are left-truncated at the time specified in enter.

Using stset to declare survival characteristics

If you are familiar with Stata’s other commands for survival analysis, such as streg, you may have been surprised to see that we did not need to stset our data to specify the failure time, censoring, and the truncation variables before using the gsem command above. Most survival analysis commands rely on stset to record information on censoring and truncation.

stset can also be used to transform time in an analysis-time metric. Analysis time is the time a subject is at risk. In this metric, a time of 0 is the time when the subject becomes at risk. gsem assumes that the dependent variable is already recorded in analysis time. If you have data in another scale, such as calendar time, you will need to transform your variables.

Although not required by gsem, stset provides a convenient way to transform data into analysis time. You can also specify truncation and censoring variables just as you would before fitting survival models with other commands. See [ST] stset for details on declaring survival data using this command.

In the dataset described above, dox records the date an individual is diagnosed with coronary heart disease, cancer, or another disease of interest. fail has a nonzero code for individuals diagnosed with a disease and a zero for individuals who were censored. dob records date of birth, and doe is the date of entry to the study. We could transform the data into analysis time using stset as follows:

```
  . stset dox, failure(fail) origin(time dob) enter(time doe) id(id)
```

This syntax gives us analysis time in days. Instead, we want to express the analysis time in years, so we type

```
  . stset dox, failure(fail) origin(time dob) enter(time doe) id(id) scale(365.25)
```

| 337 | total observations |
| 0   | exclusions        |
| 337 | observations remaining, representing |
| 337 | subjects          |
| 80  | 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 `stset` command generates the variables `_t0`, `_t`, `_d`, and `_st`.

Variable `_st` is equal to 1 unless there is a problem in the settings (for example, somebody dies before being born), in which case it is equal to 0.

Variable `_t0` indicates when the individuals enter the study, in the analysis-time scale. Variable `_t` indicates when the individual failed or was censored, also in analysis-time scale. Variable `_d` is the failure indicator.

For example,

```stata
.list dob doe dox fail _t0 _t _d _st if id == 1
```

<table>
<thead>
<tr>
<th></th>
<th>dob</th>
<th>doe</th>
<th>dox</th>
<th>fail</th>
<th>_t0</th>
<th>_t</th>
<th>_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>dob</td>
<td>04jan1915</td>
<td></td>
<td></td>
<td></td>
<td>49.615332</td>
<td>61.908282</td>
<td>0</td>
</tr>
<tr>
<td>doe</td>
<td>16aug1964</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dox</td>
<td>01dec1976</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fail</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_st</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This participant was born in 1915, entered the study in 1964, and was censored in 1976. In analysis time, this is expressed as follows: The person entered the study at age 49.6 and was censored at age 61.9.

### Fitting the loglogistic model

We can use the variables created by `stset` to fit the model

```
energy
```

```
job
```

```
height
```

```
weight
```

```
logistic
```
We specify these variables directly in our `gsem` command.

```
    . gsem (_t <- energy job height weight,>
        > family(loglogistic, failure(_d) ltruncated(_t0)))
```

(Iteration log omitted)

Generalized structural equation model                                   Number of obs  =    332
Response : _t               No. of failures =       78
Family : loglogistic         Time at risk  =   4533.5715
Form : accelerated failure-time
Link : log                   Log likelihood =  -378.82795

|                | Coef. | Std. Err. |     z  |   P>|z| |      [95% Conf. Interval] |
|----------------|-------|-----------|--------|--------|--------------------------|
| _t             |       |           |        |        |                           |
| energy         | 0.0694548 | 0.0583516 | 1.19   | 0.234  | -0.0449123 - 0.1838219   |
| job            | 0.0102962 | 0.0297246 | 0.35   | 0.729  | -0.0479629 - 0.0685552   |
| height         | 0.0107453 | 0.0047026 | 2.28   | 0.022  | 0.0015283 - 0.0199623    |
| weight         | 0.0004624 | 0.0025579 | 0.18   | 0.857  | -0.004551 - 0.0054758    |
| _cons          | 2.210313  | 0.7830543 | 2.82   | 0.005  | 0.675544 - 3.745071      |

| /_t            |       |           |        |        |                           |
| logs           | -1.818008 | 0.1701509 | -2.151498 | -1.484518 |                       |

This model is fit in the accelerated failure-time metric, and `gsem` reports coefficients. We can use `estat eform` to obtain exponentiated coefficients, which are interpreted as time ratios.

```
    . estat eform
```

|                | exp(b) | Std. Err. |     z  |   P>|z| |      [95% Conf. Interval] |
|----------------|-------|-----------|--------|--------|--------------------------|
| _t             |       |           |        |        |                           |
| energy         | 1.071924 | 0.0625485 | 1.19   | 0.234  | 0.9560813 - 1.201802     |
| job            | 1.010349 | 0.0300322 | 0.35   | 0.729  | 0.9531691 - 1.07096      |
| height         | 1.010803 | 0.0047534 | 2.28   | 0.022  | 1.00153 - 1.020163       |
| weight         | 1.000463 | 0.0025591 | 0.18   | 0.857  | 0.9954593 - 1.005491     |
| _cons          | 9.118567 | 7.140333 | 2.82   | 0.005  | 1.965122 - 42.31201      |

Each of the time ratios is just above 1, so an increase in any of the covariates would slightly increase the expected time to failure (for example, having a job increases the expected time until developing a disease by 1.01). However, only the time ratio for height is significantly different from 1.

**Fitting the model with the Builder**

Use the diagram in *Fitting the loglogistic model* above for reference.

1. Open the dataset.

   In the Command window, type

   ```
   . use https://www.stata-press.com/data/r16/gsem_diet
   . stset dox, failure(fail) origin(time dob) enter(time doe) id(id) scale(365.25)
   ```

2. Open a new Builder diagram.

   Select menu item *Statistics > SEM (structural equation modeling) > Model building and estimation*.
3. Put the Builder in \texttt{gsem} mode by clicking on the $\mathcal{G}_n$ button.

4. Create the loglogistic regression component for $-t$.

Select the Add regression component tool, $\mathcal{B}^\circ$, and then click in the diagram about one-third of the way in from the left and halfway down.

In the resulting dialog box,
\begin{enumerate}
\item select $-t$ in the \textit{Dependent variable} control;
\item check \textit{Make response generalized};
\item select Loglogistic, Log in the \textit{Family/Link} control;
\item select the \textit{Select variables} radio button (it may already be selected);
\item use the \textit{Independent variables} control to select the variables energy, job, height, and weight;
\item select Left in the \textit{Independent variables' direction} control;
\item click on \texttt{OK}.
\end{enumerate}

If you wish, move the component by clicking on any variable and dragging it.

5. Specify censoring and truncation variables.

\begin{enumerate}
\item Choose the Select tool, \texttt{\textcolor{red}{h}}.
\item Click on the box for $-t$.
\item In the Contextual Toolbar, click on the \texttt{Properties...} button.
\item In the resulting \textit{Variable properties} dialog box, click on the \textit{Failure and truncation...} button in the \textit{Variable} tab.
\item In the resulting \textit{Failure and truncation} dialog box, select $-_d$ in the \textit{Failure variable} control. Check the \textit{Survival time is left-truncated} box. Select the \textit{Variable containing truncation values} radio button, and select $-_t0$ in the \textit{Variable} control. Click on \texttt{OK}.
\item Click on \texttt{OK} in the \textit{Variables properties} dialog box.
\end{enumerate}

6. Clean up.

The box for $-t$ is created closer to the independent variables than it is in the example diagram. Use the Select tool, \texttt{\textcolor{red}{h}}, and click on the box for $-t$. Drag it to the right to allow more space for results along the paths.

7. Estimate.

Click on the \texttt{Estimate} button, \texttt{\textcolor{red}{p}}, in the Standard Toolbar, and then click on \texttt{OK} in the resulting \textit{GSEM estimation options} dialog box.

You can open a completed diagram in the Builder by typing
\begin{quote}
webgetsem gsem_llog
\end{quote}

\textbf{Reference}

Example 48g — Loglogistic survival model with censored and truncated data

Also see

- [SEM] Example 47g — Exponential survival model
- [SEM] Example 49g — Multiple-group Weibull survival model
- [SEM] Intro 5 — Tour of models
- [SEM] gsem — Generalized structural equation model estimation command
- [SEM] estat eform — Display exponentiated coefficients