

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/r19/gsem_diet
(Diet data with dates)

. describe

Contains data from https://www.stata-press.com/data/r19/gsem_diet.dta
Observations:      337      Diet data with dates
Variables:         11      16 Jan 2025 11:24
                        (_dta has notes)
```

| Variable name | Storage type | Display format | Value label | Variable 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 (1000kcal/day) |
| height | float | %9.0g | | Height (cm) |
| weight | float | %9.0g | | Weight (kg) |
| hienergy | byte | %9.0g | | Indicator for high energy |
| doe | int | %td | | Date of entry |
| dox | int | %td | | Date of exit |
| dob | int | %td | | Date of birth |

```
Sorted by: id

. notes

_dta:
  1. Data from Morris, J. N., J. W. Marr, and D. G. Clayton. 1977. Diet and
     heart: A postscript. British Medical Journal 19: 1307–1314.
     https://doi.org/10.1136/bmj.2.6098.1307.
```

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)
Survival-time data settings
      ID variable: id
      Failure event: fail!=0 & fail<.
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 | |
| 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,

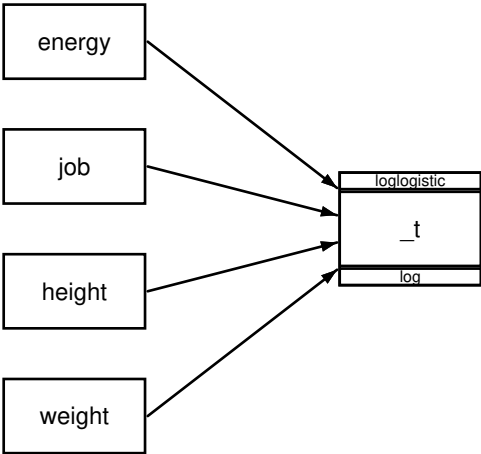
```
. list dob doe dox fail _t0 _t _d _st if id == 1
```

| | | | | | | | | |
|----|-----------|-----------|-----------|------|-----------|-----------|----|-----|
| 1. | dob | doe | dox | fail | _t0 | _t | _d | _st |
| | 04jan1915 | 16aug1964 | 01dec1976 | 0 | 49.615332 | 61.908282 | 0 | 1 |

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



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
Response: _t
Family: Log-logistic
Form: Accelerated failure time
Link: Log
Log likelihood = -378.82795

Number of obs = 332
No. of failures = 78
Time at risk = 4,533.57

| | | Coefficient | Std. err. | z | P> z | [95% conf. interval] | |
|-----|--------|-------------|-----------|------|-------|----------------------|-----------|
| _t | energy | .0694548 | .0583516 | 1.19 | 0.234 | -.0449123 | .1838219 |
| | job | .0102962 | .0297246 | 0.35 | 0.729 | -.0479629 | .0685552 |
| | height | .0107453 | .0047026 | 2.28 | 0.022 | .0015283 | .0199623 |
| | weight | .0004624 | .0025579 | 0.18 | 0.857 | -.004551 | .0054758 |
| | _cons | 2.210313 | .7830543 | 2.82 | 0.005 | .6755544 | 3.745071 |
| | | | | | | | |
| /_t | | | | | | | |
| | logs | -1.818008 | .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 | .0625485 | 1.19 | 0.234 | .9560813 | 1.201802 |
| | job | 1.010349 | .0300322 | 0.35 | 0.729 | .9531691 | 1.07096 |
| | height | 1.010803 | .0047534 | 2.28 | 0.022 | 1.00153 | 1.020163 |
| | weight | 1.000463 | .0025591 | 0.18 | 0.857 | .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/r19/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 gsem mode by clicking on the  button.
4. Create the loglogistic regression component for _t.


Select the Add regression component tool, , and then click in the diagram about one-third of the way in from the left and halfway down.

In the resulting dialog box,


- a. select _t in the *Dependent variable* control;
- b. check *Make response generalized*;
- c. select Loglogistic, Log in the *Family/Link* control;
- d. select the *Select variables* radio button (it may already be selected);
- e. use the *Independent variables* control to select the variables energy, job, height, and weight;
- f. select Left in the *Independent variables' direction* control;
- g. click on **OK**.

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


5. Specify censoring and truncation variables.

- Choose the Select tool, .
- Click on the box for `_t`.
- In the Contextual Toolbar, click on the **Properties...** button.
- In the resulting *Variable properties* dialog box, click on the **Failure and truncation...** button in the **Variable** tab.
- In the resulting *Failure and truncation* dialog box, select `_d` in the *Failure variable* control. Check the *Survival time is left-truncated* box. Select the *Variable containing truncation values* radio button, and select `_t0` in the *Variable* control. Click on **OK**.
- Click on **OK** in the *Variables properties* dialog box.

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, , 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 **Estimate** button, , in the Standard Toolbar, and then click on **OK** in the resulting *GSEM estimation options* dialog box.

You can open a completed diagram in the Builder by typing

```
. webgetsem gsem_llog
```

Reference

Morris, J. N., J. W. Marr, and D. G. Clayton. 1977. Diet and heart: A postscript. *BMJ* 19: 1307–1314. <https://doi.org/10.1136/bmj.2.6098.1307>.

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

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).