Title

stata.com

stintreg — Parametric models for interval-censored survival-time data

Description

`stintreg` fits parametric models to survival-time data that can be uncensored, right-censored, left-censored, or interval-censored. These models are generalizations of the models fit by `streg` to support interval-censored data. The supported survival models are exponential, Weibull, Gompertz, lognormal, loglogistic, and generalized gamma. Proportional-hazards (PH) and accelerated failure-time (AFT) parameterizations are provided.

With interval-censored data, the survival-time variables are specified with the `stintreg` command instead of using `stset`. Any `st` settings are ignored by `stintreg`.

Quick start

Weibull survival model with covariates `x1` and `x2` fit to interval-censored survival-time data with lower and upper endpoints `t1` and `t2`

```
stintreg x1 x2, interval(t1 t2) distribution(weibull)
```

Use AFT metric instead of PH metric

```
stintreg x1 x2, interval(t1 t2) distribution(weibull) time
```

Different intercepts and ancillary parameters for strata identified by `svar`

```
stintreg x1 x2, interval(t1 t2) distribution(weibull) strata(svar)
```

Lognormal survival model

```
stintreg x1 x2, interval(t1 t2) distribution(lognormal)
```

As above, but also model the logarithm of ancillary parameter as the linear combination of covariates `z1` and `z2`

```
stintreg x1 x2, interval(t1 t2) distribution(lognormal)  ///
     ancillary(z1 z2)
```

Menu

Statistics > Survival analysis > Regression models > Interval-censored parametric survival models

Syntax

```
stintreg [indepvars] [if] [in] [weight], interval(t_l t_u) distribution(distname)
    [options]
```

options	Description
Model	
*interval(t_l t_u)	lower and upper endpoints for the censoring interval
noconstant	suppress constant term
*distribution(*distname*)	specify survival distribution
time	use accelerated failure-time metric
Model 2	
strata(*varname*)	strata ID variable
offset(*varname*)	include *varname* in model with coefficient constrained to 1
ancillary(*varlist*)	use *varlist* to model the first ancillary parameter
anc2(*varlist*)	use *varlist* to model the second ancillary parameter
constraints(*constraints*)	apply specified linear constraints
collinear	keep collinear variables
epsilon(#)	tolerance to treat observations as uncensored; default is epsilon(1e-6)
SE/Robust	
vce(*vcetype*)	*vcetype* may be oim, robust, cluster *clustvar*, opg, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)
nohr	do not report hazard ratios
tratio	report time ratios
noheader	suppress header from coefficient table
nocnsreport	do not display constraints
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
maximize_options	control the maximization process; seldom used
coeflegend	display legend instead of statistics

*interval(t_l t_u) and distribution(*distname*) are required.

distname	Description
exponential	exponential survival distribution
gompertz	Gompertz survival distribution
loglogistic	loglogistic survival distribution
llogistic	synonym for loglogistic
weibull	Weibull survival distribution
lognormal	lognormal survival distribution
lnormal	synonym for lognormal
ggamma	generalized gamma survival distribution

varlist may contain factor variables; see [U] **11.4.3 Factor variables**.

bootstrap, by, fp, jackknife, nestreg, statsby, stepwise, and svy are allowed; see [U] **11.1.10 Prefix commands**.

Weights are not allowed with the bootstrap prefix; see [R] **bootstrap**.

vce() and noheader are not allowed with the svy prefix; see [SVY] **svy**.

fweights, iweights, and pweights may be specified.

coeflegend does not appear in the dialog box.

See [U] **20 Estimation and postestimation commands** for more capabilities of estimation commands.

Options

Model

interval(t_l t_u) specifies two time variables that contain the endpoints of the censoring interval. t_l represents the lower endpoint, and t_u represents the upper endpoint. interval() is required.

The interval time variables t_l and t_u should have the following form:

Type of data		t_l	t_u
uncensored data	$a = [a, a]$	a	a
interval-censored data	$(a, b]$	a	b
left-censored data	$(0, b]$	.	b
left-censored data	$(0, b]$	0	b
right-censored data	$[a, +\infty)$	a	.
missing		.	.
missing		0	.

noconstant; see [R] **estimation options**.

distribution(*distname*) specifies the survival model to be fit. distribution() is required.

time specifies that the model be fit in the accelerated failure-time metric rather than in the log relative-hazard metric or proportional hazards metric. This option is valid only for the exponential and Weibull models, because these are the only models that have both a proportional hazards and an accelerated failure-time parameterization. Regardless of metric, the likelihood function is the same, and models are equally appropriate viewed in either metric; it is just a matter of changing the interpretation.

Model 2

`strata(`*varname*`)` specifies the stratification ID variable. Observations with equal values of the variable are assumed to be in the same stratum. Stratified estimates (with equal coefficients across strata but intercepts and ancillary parameters unique to each stratum) are then obtained. *varname* may be a factor variable; see [U] **11.4.3 Factor variables**.

`offset(`*varname*`)`; see [R] **estimation options**.

`ancillary(`*varlist*`)` specifies that the ancillary parameter for the Weibull, lognormal, Gompertz, and loglogistic distributions and that the first ancillary parameter (σ) of the generalized log-gamma distribution be estimated as a linear combination of *varlist*.

When an ancillary parameter is constrained to be strictly positive, the logarithm of the ancillary parameter is modeled as a linear combination of *varlist*.

`anc2(`*varlist*`)` specifies that the second ancillary parameter (κ) for the generalized log-gamma distribution be estimated as a linear combination of *varlist*.

`constraints(`*constraints*`)`, `collinear`; see [R] **estimation options**.

`epsilon(#)` specifies that observations with $t_u - t_l < \#$ be treated as uncensored. The default is `epsilon(1e-6)`.

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`oim`, `opg`), that are robust to some kinds of misspecification (`robust`), that allow for intragroup correlation (`cluster` *clustvar*), and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [R] ***vce_option***.

Reporting

`level(#)`; see [R] **estimation options**.

`nohr`, which may be specified at estimation or upon redisplaying results, specifies that coefficients rather than exponentiated coefficients be displayed, that is, that coefficients rather than hazard ratios be displayed. This option affects only how coefficients are displayed, not how they are estimated.

This option is valid only for models with a natural proportional hazards parameterization: exponential, Weibull, and Gompertz. These three models, by default, report hazard ratios (exponentiated coefficients).

`tratio` specifies that exponentiated coefficients, which are interpreted as time ratios, be displayed. `tratio` is appropriate only for the loglogistic, lognormal, and generalized gamma models, or for the exponential and Weibull models when fit in the accelerated failure-time metric.

`tratio` may be specified at estimation or upon replay.

`noheader` suppresses the output header, either at estimation or upon replay.

`nocnsreport`; see [R] **estimation options**.

display_options: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(`*style*`)`, `cformat(%`*fmt*`)`, `pformat(%`*fmt*`)`, `sformat(%`*fmt*`)`, and `nolstretch`; see [R] **estimation options**.

Maximization

maximize_options: difficult, technique(*algorithm_spec*), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(*init_specs*); see [R] **maximize**. These options are seldom used.

Setting the optimization type to technique(bhhh) resets the default *vcetype* to vce(opg).

The following option is available with stintreg but is not shown in the dialog box:

coeflegend; see [R] **estimation options**.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Introduction

stintreg fits parametric models to survival-time data, which can be uncensored, right-censored, left-censored, or interval-censored. These models are generalizations of the models fit by streg, because they extend the censoring mechanism beyond right-censoring.

In survival analysis, we find different types of censored data. Among them, right-censored data have been studied extensively and can be analyzed using all of Stata's survival commands, including streg and stcox. Research on interval-censored data has also been popular; see, for example, Finkelstein and Wolfe (1985), Odell, Anderson, and D'Agostino (1992), Rabinowitz, Tsiatis, and Aragon (1995), Huang and Wellner (1997), Lindsey (1998), Lindsey and Ryan (1998), Sun (2006), and Sun and Li (2014). Interval censoring occurs when the failure time of interest is not exactly observed but is known only to lie within some interval (for example, Kalbfleisch and Prentice 2002). Uncensored, right-censored, and left-censored data are special cases of interval-censored data. In these cases, the interval reduces to a single point, is unbounded on the right, or is bounded by zero on the left.

Interval-censored survival-time data arise in many areas including medical, epidemiological, financial, and sociological studies. A study may lead to survival-time data with different types of censoring. Consider a medical study that involves periodic follow-ups with patients who had breast cancer. In this case, patients are tested on a regular basis, but the time to the recurrence of the cancer may not be measured exactly. If cancer recurs before the first visit, the observation is called left-censored. If cancer recurs between two visits, the observation is called interval-censored. If there is no recurrence by the last visit, the observation is right-censored. To analyze such data, you may fit parametric survival models using stintreg.

Regardless of the type of censoring, stintreg requires the survival outcome to be stored in the dataset as interval data. That is, two time variables, t_l and t_u, that contain the endpoints of the time interval must be specified in the interval() option. If the data are left-censored, the lower endpoint is zero and may be represented in t_l by either a missing value (.) or zero. If the data are right-censored, the upper endpoint is $+\infty$ and is represented in t_u by a missing value. Uncensored data are represented by the two endpoints that are equal. If $0 < t_l < t_u < \infty$, the data

are interval-censored. Truly missing values must be represented by missing values in both t_l and t_u or by a 0 in t_l and a missing value in t_u. Typing `stset` ([ST] **stset**) is unnecessary, and `stintreg` will ignore any settings of `stset` for the usual trivariate response variable (t_0, t, d). `stintreg` does not support data exhibiting delayed entry, gaps, time-varying covariates, and multiple failures.

Two often-used parametric models for adjusting survivor functions for the effects of covariates are the AFT models and the multiplicative or PH models. The survival models supported by `stintreg` are exponential, Weibull, Gompertz, lognormal, loglogistic, and generalized gamma. The lognormal, loglogistic, and generalized gamma models are implemented as AFT models. The exponential and Weibull models are implemented as both AFT and PH models, and the Gompertz model is implemented only in the PH metric. See *Remarks and examples* in [ST] **streg** for more details about the supported models and distributions.

Types of interval censoring

Interval censoring can occur in different forms, and each form represents one type of interval-censored survival-time data. `stintreg` accommodates two important types of interval-censored data that are commonly used in practice: case II interval-censored data and case I interval-censored data. Case II interval-censored data are also referred to as general interval-censored data, and case I interval-censored data are also referred to as current status data. We describe each censoring type in detail below. Also see Sun (2006) for more information about different types of interval censoring.

Case II interval-censored data

The most general case of interval censoring is case II interval censoring. This type of interval censoring occurs when we do not know the exact failure time t, but only know that the failure happened within a random time interval $(t_l, t_u]$, or before the right endpoint of the time interval t_u, or after the left endpoint of the time interval t_l. The following is an example of case II interval-censored data, which contain left-, right-, and interval-censored observations.

▷ Example 1: Case II interval censoring

Sun (2006) presented parametric analysis of a retrospective study of early breast cancer patients, originally from Finkelstein and Wolfe (1985), that compared the cosmetic effects of two cancer treatments: radiotherapy alone versus radiotherapy plus adjuvant chemotherapy. There were 46 radiotherapy-only patients and 48 radiation-plus-chemotherapy patients who were observed every four to six months. Patients had different visit times and durations between visits. At each visit, the physician recorded a measure of breast retraction. The event of interest was breast retraction. Because patients were observed at random follow-up times, the exact time of breast retraction was not observed and was known only to fall in the interval between visits. The data consist of two interval variables, `ltime` and `rtime`, that represent the last clinic visit time when breast retraction had not yet occurred and the first clinic visit time when breast retraction was detected.

To study the effect of treatment on breast retraction, we fit a Weibull model of time to breast retraction on treatment `treat` using `stintreg`. Unlike `streg`, in which the survival variables are set using `stset` and do not appear in the command, the interval variables `ltime` and `rtime` are required for `stintreg` and are specified in the `interval()` option:

```
. use http://www.stata-press.com/data/r15/cosmesis
(Cosmetic Deterioration of Breast Cancer Patients)
. stintreg i.treat, interval(ltime rtime) distribution(weibull)
Fitting constant-only model:
Iteration 0:   log likelihood = -200.17506
Iteration 1:   log likelihood = -175.09602
Iteration 2:   log likelihood = -153.99615
Iteration 3:   log likelihood = -148.93441
Iteration 4:   log likelihood = -148.68479
Iteration 5:   log likelihood = -148.65596
Iteration 6:   log likelihood = -148.65584
Iteration 7:   log likelihood = -148.65584
Fitting full model:
Iteration 0:   log likelihood = -148.65584
Iteration 1:   log likelihood = -143.53903
Iteration 2:   log likelihood =  -143.1932
Iteration 3:   log likelihood = -143.19228
Iteration 4:   log likelihood = -143.19228
Weibull PH regression                           Number of obs     =         94
                                                  Uncensored      =          0
                                                  Left-censored   =          5
                                                  Right-censored  =         38
                                                  Interval-cens.  =         51
                                                LR chi2(1)        =      10.93
Log likelihood = -143.19228                     Prob > chi2       =     0.0009
------------------------------------------------------------------------------
             |  Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |
Radio+Chemo  |   2.498526   .7069467     3.24   0.001     1.434961    4.350383
       _cons |   .0018503   .0013452    -8.66   0.000      .000445     .007693
-------------+----------------------------------------------------------------
       /ln_p |   .4785787   .1198973     3.99   0.000     .2435843     .713573
-------------+----------------------------------------------------------------
           p |   1.613779   .1934877                      1.275814    2.041272
         1/p |   .6196635    .074296                      .4898907    .7838134
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline hazard.
```

The header above the coefficient table summarizes censored and uncensored observations. There are 38 patients who did not experience breast retraction by the last visit, resulting in right-censored observations. There are 5 patients who had breast retraction before their first follow-up, resulting in left-censored observations. There are no uncensored observations, so the remaining 51 observations are interval-censored.

By default, the hazard ratios are reported instead of the natural coefficients. The estimated hazard ratio of the radiotherapy plus chemotherapy is approximately 2.5 with a 95% confidence interval of $[1.435, 4.350]$, indicating significantly higher risk to develop breast retraction using this treatment than radiotherapy only. In other words, the adjuvant chemotherapy increases the risk of breast retraction. The shape parameter is estimated as $\ln(p)$, but p and $1/p = \sigma$ are also reported. The estimated p is greater than 1, indicating that the hazard of breast retraction increases with time.

By default, `stintreg` uses the PH parameterization for the Weibull model, but we can specify the `time` option to request the AFT parameterization.

```
. stintreg i.treat, interval(ltime rtime) distribution(weibull) time
  (iteration log omitted)
Weibull AFT regression                          Number of obs     =         94
                                                  Uncensored      =          0
                                                  Left-censored   =          5
                                                  Right-censored  =         38
                                                  Interval-cens.  =         51
                                                LR chi2(1)        =      10.93
Log likelihood = -143.19228                     Prob > chi2       =     0.0009
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
treat						
Radio+Chemo	-.5674261	.175814	-3.23	0.001	-.9120151	-.2228371
_cons	3.899163	.1405986	27.73	0.000	3.623595	4.174731
/ln_p	.4785789	.119897	3.99	0.000	.2435851	.7135726
p	1.613779	.1934873			1.275815	2.041271
1/p	.6196634	.0742958			.4898909	.7838128

With the AFT parameterization, coefficients are reported by default, but we can use the `tratio` option to display time ratios.

◁

▷ Example 2: Comparing distributions

To compare different models, let's fit the model from example 1 but use the generalized gamma distribution instead. The hazard function of the generalized gamma distribution is extremely flexible, allowing for many different shapes. Weibull, exponential, and lognormal distributions are all special cases of the generalized gamma distribution. Therefore, we can use the generalized gamma model to evaluate and select an appropriate parametric model for the data. When $\kappa = 0$, the generalized gamma model reduces to the lognormal model. When $\kappa = 1$, the generalized gamma model reduces to the Weibull model.

```
. stintreg i.treat, interval(ltime rtime) distribution(ggamma)
  (iteration log omitted)
Generalized gamma AFT regression                Number of obs     =         94
                                                  Uncensored      =          0
                                                  Left-censored   =          5
                                                  Right-censored  =         38
                                                  Interval-cens.  =         51
                                                LR chi2(1)        =      11.26
Log likelihood = -142.71767                     Prob > chi2       =     0.0008
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
treat						
Radio+Chemo	-.5696387	.1686355	-3.38	0.001	-.9001581	-.2391192
_cons	4.009316	.1721275	23.29	0.000	3.671952	4.346679
/lnsigma	-.7016456	.2793936	-2.51	0.012	-1.249247	-.1540442
/kappa	1.532208	.6176603	2.48	0.013	.3216162	2.7428
sigma	.4957688	.1385146			.2867206	.8572342

The Wald test of H_0: $\kappa = 0$ is reported in the output above. The p-value is 0.013, indicating that the lognormal model is not appropriate. We can test the hypothesis that $\kappa = 1$ using the `test` command:

```
. test /kappa = 1
 ( 1)  [/]kappa = 1
           chi2(  1) =    0.74
         Prob > chi2 =    0.3889
```

The above Wald test of H_0: $\kappa = 1$ has a p-value of 0.39, suggesting that the Weibull model may be appropriate for these data.

◁

Case I interval-censored data

Case I interval-censored data arise when the only survival information available is whether the event of interest occurred before or after the observed time, leading to data in which an observation is either left-censored or right-censored. As such, case I interval-censored data can be viewed as a special case of case II interval-censored data without uncensored and interval-censored on $(a, b]$ observations. Case I interval-censored data occur when subjects are observed only once, and thus we can only know whether the event had already happened before we observed them. Such data are common in demographical studies, where they are also known as current status data. In addition to demographical studies, case I interval-censored data occur in other fields including epidemiological studies, cross-sectional studies, and tumorigenicity experiments. See Huang and Wellner (1997) and Sun (2006) for more information.

The `stintreg` command requires that case I interval-censored data are recorded by two interval time variables that identify which observations are left-censored and which observations are right-censored.

Example 3: Case I interval censoring

We consider the data from Hoel and Walburg (1972) on nonlethal lung tumors for 144 male mice. The death time of each mouse (`death`) and an indicator of whether the lung tumor was present by the time of death (`status`) were reported. The type of environment (`group`) in which those mice lived, either conventional environment (`CE`) or germ-free environment (`GE`), was also reported. The goal of this study was to test whether different types of environment had influence on the time of tumor onset for those mice. The lung tumor was known to be nonlethal for the mice. Therefore, the tumor onset time could not be directly observed. The only available information was the observed death time and whether or not the tumor was detected at the time of death.

```
. use http://www.stata-press.com/data/r15/lungtumor
(Lung Tumor Data For Mice)
. table group status
```

Environment	Tumor status	
	No tumor	With tumor
CE	69	27
GE	13	35

```
. list in 26/30
```

	group	status	death
26.	CE	With tumor	811
27.	CE	With tumor	839
28.	CE	No tumor	45
29.	CE	No tumor	198
30.	CE	No tumor	215

Case I interval-censored data are often stored as shown above: each subject has one variable that represents the observation time and one variable that represents the status of the event of interest. To use `stintreg`, we must create two time variables to contain the lower and upper endpoints of the intervals. Because case I interval-censored data are either left-censored or right-censored, we first create two new variables, `ltime` and `rtime`, that are both equal to the observation time, `death`, then replace the lower endpoint, `ltime`, with a missing value if the tumor was detected (`status` $= 1$) and replace the upper endpoint, `rtime`, with a missing value if the tumor was not detected (`status` $= 0$).

```
. generate ltime = death
. generate rtime = death
. replace ltime = . if status == 1
(62 real changes made, 62 to missing)
. replace rtime = . if status == 0
(82 real changes made, 82 to missing)
. list in 26/30
```

	group	status	death	ltime	rtime
26.	CE	With tumor	811	.	811
27.	CE	With tumor	839	.	839
28.	CE	No tumor	45	45	.
29.	CE	No tumor	198	198	.
30.	CE	No tumor	215	215	.

Now, we fit the model using the exponential distribution.

```
. stintreg i.group, interval(ltime rtime) distribution(exponential)
  (iteration log omitted)
Exponential PH regression                       Number of obs     =        144
                                                  Uncensored      =          0
                                                  Left-censored   =         62
                                                  Right-censored  =         82
                                                  Interval-cens.  =          0
                                                LR chi2(1)        =      16.09
Log likelihood = -81.325875                     Prob > chi2       =     0.0001
------------------------------------------------------------------------------
             | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |
          GE |    2.90202   .7728318     4.00   0.000     1.721942    4.890828
       _cons |   .0005664   .0001096   -38.63   0.000     .0003876    .0008277
------------------------------------------------------------------------------
Note: _cons estimates baseline hazard.
```

The estimated hazard for the mice in the germ-free environment is approximately three times the hazard for those in the conventional environment. In other words, the mice in the germ-free environment had higher lung tumor incidence than those in the conventional environment.

◁

Parameterization of ancillary parameters

`stintreg`'s `ancillary()` and `anc2()` options allow us to parameterize ancillary parameters in terms of covariates. By default, all ancillary parameters are estimated as being constant. By specifying, for example,

```
. stintreg x1 x2, interval(ltime rtime) distribution(weibull) ancillary(z1 z2)
```

the logarithm of the ancillary parameter p is modeled using the linear predictor of `z1` and `z2`. The `anc2()` option models the second ancillary parameter κ for the generalized log-gamma distribution.

▷ Example 4: Modeling the ancillary parameters

Consider the data described in table 2.3 of Sun (2006) (originally from Richman, Grimes, and Lagakos 1990) on times to resistance to the drug zidovudine for AIDS patients. Covariates of interest are the stage of the disease, `stage` (0 = early stage, 1 = late stage) and the dose level of the treatment, `dose` (0 = low dose, 1 = high dose). The time intervals, in months, are stored in variables `ltime` and `rtime`.

To investigate whether `stage` has any effect on time to drug resistance, we fit the Weibull model using `stintreg`. To later compare results with another model that reports coefficients, we use the `nohr` option here to display the untransformed coefficients.

```
. use http://www.stata-press.com/data/r15/aids, clear
(Time to Zidovudine Resistance)
```

```
. stintreg i.stage, interval(ltime rtime) distribution(weibull) nohr
  (iteration log omitted)
Weibull PH regression                           Number of obs     =         31
                                                  Uncensored      =          0
                                                  Left-censored   =         15
                                                  Right-censored  =         13
                                                  Interval-cens.  =          3
                                                LR chi2(1)        =      10.02
Log likelihood =  -13.27946                     Prob > chi2       =     0.0016
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.stage	1.910652	.6604417	2.89	0.004	.6162106	3.205094
_cons	-7.952872	3.000565	-2.65	0.008	-13.83387	-2.071873
/ln_p	1.036663	.3978289	2.61	0.009	.2569325	1.816393
p	2.819791	1.121795			1.292958	6.149638
1/p	.3546362	.1410845			.1626112	.7734204

Out of the 31 patients, 13 patients are right-censored, 15 patients are left-censored, and only 3 patients are interval-censored. The estimated coefficient for patients in their late stage of the disease is 1.91; their hazard of resisting zidovudine is approximately $\exp(1.91) = 6.75$ times the hazard for patients in their early stage.

Suppose we believe that the hazards for different dose levels have different shape parameters. We can accommodate this by specifying the `ancillary()` option.

```
. stintreg i.stage, interval(ltime rtime) distribution(weibull) ancillary(i.dose)
note: option nohr is implied if option strata() or ancillary() is specified
  (iteration log omitted)
Weibull PH regression                           Number of obs     =         31
                                                  Uncensored      =          0
                                                  Left-censored   =         15
                                                  Right-censored  =         13
                                                  Interval-cens.  =          3
                                                LR chi2(1)        =      12.20
Log likelihood = -11.214877                     Prob > chi2       =     0.0005
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ltime						
1.stage	2.795073	1.167501	2.39	0.017	.5068139	5.083332
_cons	-10.8462	4.233065	-2.56	0.010	-19.14286	-2.549547
ln_p						
1.dose	.1655302	.0874501	1.89	0.058	-.0058689	.3369292
_cons	1.252361	.4143257	3.02	0.003	.4402972	2.064424

With the `ancillary()` option, results are displayed as coefficients by default; see the technical note below. From the above results, $\widehat{\ln(p)}_{\text{low}} = 1.25$ for patients with low dose and $\widehat{\ln(p)}_{\text{high}} = 1.25 + 0.17 = 1.42$ for patients with high dose. Thus, $\widehat{p}_{\text{low}} = 3.49$ and $\widehat{p}_{\text{high}} = 4.14$. When we combine this with the main equation in the model, the estimated hazards are

$$\widehat{h}(t_j|\mathbf{x}_j) = \begin{cases} 3.49 \times t_j^{3.49-1} \times \exp\left(-10.85 + 2.80 \times \texttt{stage}_j\right) & \text{if } \texttt{dose} = 0 \\ 4.14 \times t_j^{4.14-1} \times \exp\left(-10.85 + 2.80 \times \texttt{stage}_j\right) & \text{if } \texttt{dose} = 1 \end{cases}$$

◁

❑ Technical note

When fitting PH models, stintreg, by default, displays hazard ratios. If the strata() option or the ancillary() option (as in our previous example) is specified, stintreg reports coefficients instead. If either of these options is specified, ancillary parameters are no longer constant and are modeled as a function of covariates specified in those options. If any of the covariates from the main equation are used to model ancillary parameters, hazard ratios lose their interpretation. As a precaution, stintreg always displays results as coefficients when those options are used. If we want to compare results with PH models with constant ancillary parameters, we can use the nohr option to display coefficients.

The above argument also applies to time ratios when fitting AFT models. For this reason, the tratio option is not allowed with AFT models whenever strata(), ancillary(), or anc2() is specified.

❑

Stratified estimation

We can fit a stratified model by specifying the strata(*varname*) option. A stratified model means that the coefficients on the covariates are the same across strata, but the intercept and ancillary parameters are allowed to vary for each level of the strata variable.

▷ Example 5: Fitting a stratified model

Continuing with example 4, suppose that we believe that dose affects both the scale and shape of the hazard, and the effect of stage is the same for each level of dose. We refit the Weibull model, but now we also stratify on dose:

```
. stintreg i.stage, interval(ltime rtime) distribution(weibull) strata(dose)
note: option nohr is implied if option strata() or ancillary() is specified
  (iteration log omitted)
Weibull PH regression                           Number of obs     =         31
                                                  Uncensored      =          0
                                                  Left-censored   =         15
                                                  Right-censored  =         13
                                                  Interval-cens.  =          3
                                                LR chi2(2)        =      12.40
Log likelihood = -11.115197                     Prob > chi2       =     0.0020
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ltime						
1.stage	2.711532	1.084146	2.50	0.012	.5866456	4.836419
1.dose	-2.661872	5.883967	-0.45	0.651	-14.19424	8.870492
_cons	-9.143003	4.930789	-1.85	0.064	-18.80717	.5211664
ln_p						
1.dose	.453894	.670098	0.68	0.498	-.8594739	1.767262
_cons	1.051935	.6190537	1.70	0.089	-.1613879	2.265258

The indicator for level 1 of dose is added to the main equation and to the ancillary equation; level 0 is the baseline and is modeled by the constant terms.

Note that the specification above is the same as fitting the following model:

```
stintreg i.stage i.dose, interval(ltime rtime) distribution(weibull) ///
     ancillary(i.dose)
```

◁

By using `ancillary()` or `strata()`, we may fit a wide variety of models; see *Stratified estimation* in [ST] **streg** for details. These models may be compared using Wald or likelihood-ratio tests when the models in question are nested or by using the AIC for nonnested models. Modeling of ancillary parameters and stratification is also available for AFT models.

Stored results

`stintreg` stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(N_unc)`	number of uncensored observations
`e(N_lc)`	number of left-censored observations
`e(N_rc)`	number of right-censored observations
`e(N_int)`	number of interval-censored observations
`e(k)`	number of parameters
`e(k_eq)`	number of equations in `e(b)`
`e(k_eq_model)`	number of equations in overall model test
`e(k_aux)`	number of auxiliary parameters
`e(k_dv)`	number of dependent variables
`e(df_m)`	model degrees of freedom
`e(ll)`	log likelihood
`e(ll_0)`	log likelihood, constant-only model
`e(N_clust)`	number of clusters
`e(chi2)`	χ^2
`e(aux_p)`	ancillary parameter (`weibull`)
`e(gamma)`	ancillary parameter (`gompertz`, `loglogistic`)
`e(sigma)`	ancillary parameter (`ggamma`, `lnormal`)
`e(kappa)`	ancillary parameter (`ggamma`)
`e(epsilon)`	tolerance for uncensored observations
`e(p)`	significance
`e(rank)`	rank of `e(V)`
`e(rank0)`	rank of `e(V)`, constant-only model
`e(ic)`	number of iterations
`e(rc)`	return code
`e(converged)`	`1` if converged, `0` otherwise

Macros

`e(cmd)`	model or regression name
`e(cmd2)`	`stintreg`
`e(cmdline)`	command as typed
`e(depvar)`	names of time interval variables specified in `interval()`
`e(distribution)`	distribution
`e(strata)`	stratum variable
`e(title)`	title in estimation output
`e(clustvar)`	name of cluster variable
`e(wtype)`	weight type
`e(wexp)`	weight expression
`e(vce)`	*vcetype* specified in `vce()`
`e(vcetype)`	title used to label Std. Err.
`e(frm2)`	`hazard` or `time`
`e(chi2type)`	`Wald` or `LR`; type of model χ^2 test
`e(offset1)`	offset for main equation
`e(opt)`	type of optimization

e(which)	max or min; whether optimizer is to perform maximization or minimization
e(ml_method)	type of ml method
e(user)	name of likelihood-evaluator program
e(technique)	maximization technique
e(properties)	b V
e(predict)	program used to implement predict
e(predict_sub)	predict subprogram
e(marginsok)	predictions allowed by margins
e(marginsnotok)	predictions disallowed by margins
e(asbalanced)	factor variables fvset as asbalanced
e(asobserved)	factor variables fvset as asobserved
Matrices	
e(b)	coefficient vector
e(Cns)	constraints matrix
e(ilog)	iteration log (up to 20 iterations)
e(gradient)	gradient vector
e(V)	variance–covariance matrix of the estimators
e(V_modelbased)	model-based variance
Functions	
e(sample)	marks estimation sample

Methods and formulas

Methods and formulas are presented under the following headings:

Introduction

Consider survival-time data that consists of n independent observations. Let t_j represent the survival time for the event of interest for observation j, $j = 1, \ldots, n$.

For a given survivor function, $S(t)$, the density function is obtained as

$$f(t) = -\frac{d}{dt}S(t)$$

and the hazard function (the instantaneous rate of failure) is obtained as

$$h(t) = \frac{f(t)}{S(t)} = -\frac{\log S(t)}{dt}$$

Let $\mathbf{x}_j$ denote a vector of covariates for observation j, and let $\boldsymbol{\beta}$ denote a vector of regression coefficients. Let $S_j(t) = S(t|\mathbf{x} = \mathbf{x}_j)$ be the covariate-adjusted survivor function and similarly define $h_j(t)$ and $f_j(t)$.

stintreg supports six survival distributions: exponential, Weibull, Gompertz, lognormal, loglogistic, and generalized gamma; and two parameterizations for the effects of covariates: PH and AFT. The parameterization and ancillary parameters for each distribution are summarized in table 1 below.

Distributions and parameterizations

The PH model assumes that the hazard function has the form

$$h_j(t) = h_0(t)\exp(\mathbf{x}_j\boldsymbol{\beta})$$

for some baseline hazard function $h_0(t)$. For the `stintreg` command, $h_0(t)$ is assumed to be parametric and the supported distributions are exponential, Weibull, and Gompertz. This model specifies that the covariates have a multiplicative effect on the hazard function. The covariate-adjusted survivor function $S_j(t)$ is obtained as

$$S_j(t) = \{S_0(t)\}^{\exp(\mathbf{x}_j\boldsymbol{\beta})}$$

where the baseline survivor function $S_0(t) = \exp\{-\int_0^t h_0(s)ds\}$.

In the AFT model, the natural logarithm of the survival time, $\log t$, is expressed as a linear function of the covariates, yielding the linear model

$$\log t_j = \mathbf{x}_j\boldsymbol{\beta} + z_j$$

where z_j is the error term with density $f()$. The distributional form of the error term determines the regression model. If we let $f()$ be the normal density, the lognormal regression model for t_j is obtained. Similarly, by letting $f()$ be the logistic density, the loglogistic regression is obtained. Setting $f()$ equal to the extreme-value density yields the exponential and the Weibull regression models. The effect of covariates is also multiplicative, but on time t_j, by a factor of $\exp(-\mathbf{x}_j\boldsymbol{\beta})$. Depending on whether this factor is greater or less than one, time is either accelerated or decelerated.

Table 1 below describes the supported survival models, their parameterizations, and the corresponding ancillary parameters.

Table 1. Parametric survival distributions supported by `stintreg`

Distribution	Metric	Survivor function	Parameterization	Ancillary parameters
Exponential	PH	$\exp(-\lambda_j t_j)$	$\lambda_j = \exp(\mathbf{x}_j\boldsymbol{\beta})$	
Exponential	AFT	$\exp(-\lambda_j t_j)$	$\lambda_j = \exp(-\mathbf{x}_j\boldsymbol{\beta})$	
Weibull	PH	$\exp(-\lambda_j t_j^p)$	$\lambda_j = \exp(\mathbf{x}_j\boldsymbol{\beta})$	p
Weibull	AFT	$\exp(-\lambda_j t_j^p)$	$\lambda_j = \exp(-p\mathbf{x}_j\boldsymbol{\beta})$	p
Gompertz	PH	$\exp\{-\lambda_j\gamma^{-1}(e^{\gamma t_j}-1)\}$	$\lambda_j = \exp(\mathbf{x}_j\boldsymbol{\beta})$	γ
Lognormal	AFT	$1-\Phi\left\{\frac{\log(t_j)-\mu_j}{\sigma}\right\}$	$\mu_j = \mathbf{x}_j\boldsymbol{\beta}$	σ
Loglogistic	AFT	$\{1+(\lambda_j t_j)^{1/\gamma}\}^{-1}$	$\lambda_j = \exp(-\mathbf{x}_j\boldsymbol{\beta})$	γ
Generalized gamma				
if $\kappa > 0$	AFT	$1-I(\gamma,u)$	$\mu_j = \mathbf{x}_j\boldsymbol{\beta}$	σ,κ
if $\kappa = 0$	AFT	$1-\Phi(z)$	$\mu_j = \mathbf{x}_j\boldsymbol{\beta}$	σ,κ
if $\kappa < 0$	AFT	$I(\gamma,u)$	$\mu_j = \mathbf{x}_j\boldsymbol{\beta}$	σ,κ

where $\Phi(z)$ is the standard normal cumulative distribution. For the generalized gamma, $\gamma = |\kappa|^{-2}$, $u = \gamma\exp(|\kappa|z)$, $I(a,x)$ is the incomplete gamma function, and $z = \text{sign}(\kappa)\{\log(t_j) - \mu_j\}/\sigma$.

Parameter estimation using interval-censored data

Suppose that t_j is not observed and that only the lower and upper endpoints of the time interval, t_{lj} and t_{uj}, where $t_j \in (t_{lj}, t_{uj}]$, are observed. `stintreg` estimates $\boldsymbol{\beta}$ and the ancillary parameters via maximum likelihood. For interval-censored observations, the log likelihood is given by

$$\log L = \sum_{j=1}^{n} \{ \log S_j(t_{lj}) - \log S_j(t_{uj}) \}$$

Implicit in the above log-likelihood expression are the regression parameters, $\boldsymbol{\beta}$, and the ancillary parameters, because both are components of the chosen $S_j(t)$; see table 1.

For case II interval-censored data, the log likelihood can be written as

$$\log L = \sum_{j \in UC} \log f_j(t_{lj}) + \sum_{j \in RC} \log S_j(t_{lj}) + \sum_{j \in LC} \{1 - \log S_j(t_{uj})\} + \sum_{j \in IC} \{ \log S_j(t_{lj}) - \log S_j(t_{uj}) \}$$

where the set UC contains uncensored observations, RC contains right-censored observations, LC contains left-censored observations, and IC contains interval-censored observations.

For case I interval-censored data, with only right-censored and left-censored observations, the log likelihood reduces to

$$\log L = \sum_{j \in RC} \log S_j(t_{lj}) + \sum_{j \in LC} \{1 - \log S_j(t_{uj})\}$$

Specifying `ancillary()`, `anc2()`, or `strata()` will parameterize the ancillary parameter(s) by using the linear predictor, $\mathbf{z}_j\boldsymbol{\alpha}_z$, where the covariates, $\mathbf{z}_j$, need not be distinct from $\mathbf{x}_j$. Here `stintreg` will report estimates of $\boldsymbol{\alpha}_z$ in addition to estimates of $\boldsymbol{\beta}$. The log likelihood here is simply the log likelihood given above, with $\mathbf{z}_j\boldsymbol{\alpha}_z$ substituted for the ancillary parameter. If the ancillary parameter is constrained to be strictly positive, its logarithm is parameterized instead; that is, we substitute the linear predictor for the logarithm of the ancillary parameter in the above log likelihood. The gamma model has two ancillary parameters, σ and κ; we parameterize σ by using `ancillary()` and κ by using `anc2()`, and the linear predictors used for each may be distinct. Specifying `strata()` includes indicator variables for the strata in the main equation, and uses them to parameterize any ancillary parameters that exist for the chosen model.

This command supports the Huber/White/sandwich estimator of the variance and its clustered version using `vce(robust)` and `vce(cluster` *clustvar*`)`, respectively. See [P] **_robust**, particularly *Maximum likelihood estimators* and *Methods and formulas*. If the assumption of independence of the observations is highly questionable, this means that the conventional estimate of variance is not appropriate. We strongly advise that you use the `vce(robust)` and `vce(cluster` *clustvar*`)` options here.

`stintreg` also supports estimation with survey data. For details on VCEs with survey data, see [SVY] **variance estimation**.

References

Finkelstein, D. M., and R. A. Wolfe. 1985. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* 41: 933–945.

Hoel, D. G., and H. E. Walburg, Jr. 1972. Statistical analysis of survival experiments. *Journal of the National Cancer Institute* 49: 361–372.

Huang, J., and J. A. Wellner. 1997. Interval censored survival data: A review of recent Progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, ed. D. Y. Lin and T. R. Fleming, 123–169. New York: Springer.

Kalbfleisch, J. D., and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York: Wiley.

Lindsey, J. C., and L. M. Ryan. 1998. Methods for interval-censored data. *Statistics in Medicine* 17: 219–238.

Lindsey, J. K. 1998. A study of interval censoring in parametric regression models. *Lifetime Data Analysis* 4: 329–354.

Odell, P. M., K. M. Anderson, and R. B. D'Agostino. 1992. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics* 48: 951–959.

Rabinowitz, D., A. A. Tsiatis, and J. Aragon. 1995. Regression with interval-censored data. *Biometrika* 82: 501–513.

Richman, D. D., J. M. Grimes, and S. W. Lagakos. 1990. Effect of stage of disease and drug dose on zidovudine susceptibilities of isolates of human immunodeficiency virus. *Journal of Acquired Immune Deficiency Syndromes* 3: 743–746.

Sun, J. 2006. *The Statistical Analysis of Interval-Censored Failure Time Data*. New York: Springer.

Sun, J., and J. Li. 2014. Interval censoring. In *Handbook of Survival Analysis*, ed. J. P. Klein, H. C. van Houwelingen, J. G. Ibrahim, and T. H. Scheike, 369–390. Boca Raton, FL: CRC Press.

Also see

[ST] **stintreg postestimation** — Postestimation tools for stintreg

[ST] **stcurve** — Plot survivor, hazard, cumulative hazard, or cumulative incidence function

[ST] **stcox** — Cox proportional hazards model

[ST] **streg** — Parametric survival models

[ME] **meintreg** — Multilevel mixed-effects interval regression

[R] **intreg** — Interval regression

[SVY] **svy estimation** — Estimation commands for survey data

[XT] **xtintreg** — Random-effects interval-data regression models

[U] **20 Estimation and postestimation commands**