**stteffects ra — Survival-time regression adjustment**

**Description**

`stteffects ra` estimates the average treatment effect (ATE), the average treatment effect on the treated (ATET), and the potential-outcome means (POMs) from observational survival-time data by regression adjustment (RA). RA uses averages of treatment-specific predicted mean survival times to estimate mean survival times for each potential outcome. Contrasts of these predicted mean survival times estimate the treatment effects. `stteffects ra` offers several choices for the model used to predict mean survival time. Binary and multivalued treatments are accommodated.

See [CAUSAL] `stteffects intro` for an overview of estimating treatment effects from observational survival-time data.

**Quick start**

Specify `time` as observed failure time and `fail` as failure indicator

```
    stset time, failure(fail)
```

ATE from a Weibull model for `time` on `x1` and `x2` with binary treatment `treat2`

```
    stteffects ra (x1 x2) (treat2)
```

Same as above, but estimate the ATET

```
    stteffects ra (x1 x2) (treat2), atet
```

Same as above, but estimate the potential-outcome means

```
    stteffects ra (x1 x2) (treat2), pomeans
```

ATE of `treat2` using a gamma model for `time`

```
    stteffects ra (x1 x2, gamma) (treat2)
```

ATE for each level of three-valued treatment `treat3`

```
    stteffects ra (x1 x2) (treat3)
```

Same as above, and specify that `treat3 = 3` is the control level using the value label “MyControl” for 3

```
    stteffects ra (x1 x2) (treat3), control("MyControl")
```

**Menu**

Statistics > Causal inference/treatment effects > Survival outcomes > Regression adjustment
**Syntax**

```
stteffects ra (omvarlist [, omoptions]) (tvar) [if] [in] [, stat options]
```

*omvarlist* specifies the variables that predict the survival-time variable in the outcome model.
*tvar* must contain integer values representing the treatment levels.

**omoptions**

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>weibull</td>
<td>Weibull; the default</td>
</tr>
<tr>
<td>exponential</td>
<td>exponential</td>
</tr>
<tr>
<td>gamma</td>
<td>two-parameter gamma</td>
</tr>
<tr>
<td>lnnormal</td>
<td>lognormal</td>
</tr>
<tr>
<td>ancillary</td>
<td>specify variables used to model ancillary parameter</td>
</tr>
<tr>
<td>noconstant</td>
<td>suppress constant from outcome model</td>
</tr>
</tbody>
</table>

**stat**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ate</td>
</tr>
<tr>
<td>ate</td>
</tr>
<tr>
<td>pomeans</td>
</tr>
</tbody>
</table>

**options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE/Robust</td>
</tr>
<tr>
<td>vce(vcetype)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>level(</em>)</td>
</tr>
<tr>
<td>aequations</td>
</tr>
<tr>
<td>noshow</td>
</tr>
<tr>
<td>display_options</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maximization</th>
</tr>
</thead>
<tbody>
<tr>
<td>maximize_options</td>
</tr>
<tr>
<td>iterinit(#)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>control(#)</td>
</tr>
<tr>
<td>tlevel(#)</td>
</tr>
<tr>
<td>coeflegend</td>
</tr>
</tbody>
</table>
You must stset your data before using stteffects; see [ST] stset.

omvarlist and avarlist may contain factor variables; see [U] 11.4.3 Factor variables.

bootstrap, by, collect, jackknife, and statsby are allowed; see [U] 11.1.10 Prefix commands.

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

fweights, iweights, and pweights may be specified using stset; see Weights under Remarks and examples in [ST] stset. However, weights may not be specified if you are using the bootstrap prefix.

coefflegend does not appear in the dialog box.

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

Options

ancillary(avarlist [, noconstant]) specifies the variables used to model the ancillary parameter.

By default, the ancillary parameter does not depend on covariates. Specifying ancillary(avarlist, noconstant) causes the constant to be suppressed in the model for the ancillary parameter.

noconstant; see [R] Estimation options.

stat is one of three statistics: ate, atet, or pomeans. ate is the default.

ate specifies that the average treatment effect be estimated.

atet specifies that the average treatment effect on the treated be estimated.

pomeans specifies that the potential-outcome means for each treatment level be estimated.

vce(vcetype) specifies the type of standard error reported, which includes types 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.

level(#); see [R] Estimation options.

aequations specifies that the results for the outcome-model or treatment-model parameters be displayed. By default, the results for these auxiliary parameters are not displayed.

noshow prevents stteffects ra from showing the key st variables. This option is rarely used because most people type stset, show or stset, noshow to permanently set whether they want to see these variables mentioned at the top of the output of every st command; see [ST] stset.

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

maximize_options: iterate(#), [no]log, and from(init_specs); see [R] Maximize. These options are seldom used.

init_specs is one of

matname [, skip copy]
# [, # ...], copy
iterinit(#) specifies the maximum number of iterations used to calculate the starting values. This option is seldom used.

control(#| label) specifies the level of tvar that is the control. The default is the first treatment level. You may specify the numeric level # (a nonnegative integer) or the label associated with the numeric level. control() may not be specified with the statistic pomeans. control() and tlevel() may not specify the same treatment level.

tlevel(#| label) specifies the level of tvar that is the treatment for the statistic atet. The default is the second treatment level. You may specify the numeric level # (a nonnegative integer) or the label associated with the numeric level. tlevel() may only be specified with statistic atet. tlevel() and control() may not specify the same treatment level.

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

coefflegend; see [R] Estimation options.

Remarks and examples

If you are not familiar with the framework for treatment-effects estimation from observational survival-time data, please see [CAUSAL] stteffects intro.

RA estimators use contrasts of the averages of treatment-specific predicted mean outcomes to estimate treatment effects. RA estimators use a two-step approach to estimating treatment effects:

1. For each treatment level, fit a model of the survival-time outcome on the same set of covariates.
2. Compute the averages of the predicted outcomes for each subject within each treatment level.

These averages estimate the potential-outcome means (POMs). Contrasts of these averages estimate the ATEs. By restricting the computations of the averages to the subset of treated subjects, we obtain estimates of the ATETs.

Here we note only a few entry points to the vast literature on RA estimators. Imbens (2004), Imbens and Wooldridge (2009), Cameron and Trivedi (2005, chap. 25), Wooldridge (2010, chap. 21), and Vittinghoff et al. (2012, chap. 9) provide excellent general introductions to estimating ATEs and to RA estimators in particular.

Like streg and other survival-time commands, stteffects ra uses the outcome variable and the failure indicator computed by, and optionally weights specified with, stset. stteffects ra is not appropriate for data with time-varying covariates, also known as multiple-record survival-time data, or for delayed-entry data.

Example 1: Estimating the ATE

Suppose we wish to study the effect of smoking on the time to a second heart attack among women aged 45–55 years. In our fictional sheart dataset, atime is the observed time in years to a second heart attack or censoring, and fail is the 0/1 indicator that a second heart attack was observed. (When fail is 1, atime records the time to the second heart attack; when fail is 0, atime records a censored observation of the time to a second heart attack.) We previously stset these data; see A quick tour of the estimators in [CAUSAL] stteffects intro.
The treatment, smoking, is stored in the 0/1 indicator `smoke`. These data also contain age at the time of the first heart attack (`age`), and indices of the level of exercise (`exercise`), diet quality (`diet`), and education (`education`) prior to the first heart attack.

We can use `stteffects ra` to estimate the ATE by RA. We model the mean survival time using the default Weibull model, controlling for `age`, `exercise`, `diet`, and `education`, and we specify that `smoke` is the treatment variable.

```
. use https://www.stata-press.com/data/r18/sheart
   (Time to second heart attack (fictional))
. stteffects ra (age exercise diet education) (smoke)
   Failure _d: fail
   Analysis time _t: atime
Iteration 0: EE criterion = 1.006e-14
Iteration 1: EE criterion = 2.305e-25
Survival treatment-effects estimation     Number of obs = 2,000
Estimator : regression adjustment
Outcome model : Weibull
Treatment model: none
Censoring model: none

|      _t | Coefficient | std. err. |      z  | P>|z|  | [95% conf. interval] |
|---------|-------------|-----------|--------|------|---------------------|
|         | Robust      |           |        |      |                     |
| ATE     |             |           |        |      |                     |
| smoke   |             |           |        |      |                     |
| (Smoker vs Nonsmoker) | -1.956657 | 0.3331787 | -5.87  | 0.000 | -2.609676 -1.303639 |
| POmean  |             |           |        |      |                     |
| smoke   |             |           |        |      |                     |
| Nonsmoker | 4.243974  | 0.2620538 | 16.20  | 0.000 | 3.730358 4.75759  |
```

When every woman smoked in the population of women aged 45–55 years who have had a heart attack, the average time to a second heart attack is estimated to be 1.96 years less than when no women in the population of interest smoked. The estimated average time to a second heart attack when no women in the population of interest smoked is 4.24 years. In other words, if every woman in the population of interest smoked, then the average time to a second heart attack would fall by an estimated 46% relative to the case when no women smoked.

Example 2: Changing the outcome model

Instead of a Weibull model for the outcome model, we could have used an exponential, a gamma, or a lognormal model. By way of comparison, we use a gamma model and the same covariates to estimate the ATE.
. stteffects ra (age exercise diet education, gamma) (smoke)

Failure _d: fail
Analysis time _t: atime
Iteration 0: EE criterion = 6.212e-25
Iteration 1: EE criterion = 7.256e-32

Survival treatment-effects estimation
Number of obs = 2,000
Estimator : regression adjustment
Outcome model : gamma
Treatment model: none
Censoring model: none

|             | Coefficient | std. err. | z     | P>|z|    | [95% conf. interval] |
|-------------|-------------|-----------|-------|--------|---------------------|
| _t          |             |           |       |        |                     |
| ATE         |             |           |       |        |                     |
| smoke       |             |           |       |        |                     |
| (Smoker vs  | -1.801787   | .2924388  | -6.16 | 0.000  | -2.374956 -1.228617 |
| Nonsmoker)  |             |           |       |        |                     |
| POmean      |             |           |       |        |                     |
| smoke       |             |           |       |        |                     |
| Nonsmoker   | 3.994327    | .2258257  | 17.69 | 0.000  | 3.551717 4.436937   |

The estimated ATE of $-1.80$ and control-level POM of 3.99 are similar to those of $-1.96$ and 4.24 obtained from the Weibull model in example 1. The ratio of the estimated ATE to the control-level POM indicates a 45% reduction instead of the 46% reduction obtained from the Weibull model.

Example 3: Estimating the ratio of the ATE to the control-level POM

The ratio of the ATE to the control-level POM measures the importance of the effect. In example 1, we computed the point estimate of this ratio from the output, but we were left without a confidence interval. In this example, we use nlcom to compute a point estimate and a confidence interval.

Below, we refit the model from example 1, specifying the coeflegend option to learn the parameter names. We use the parameter names in nlcom to estimate the ratio of the ATE to the control-level POM.
Analysis of the output shows that when every woman smoked, the average time to a second heart attack falls by an estimated 46% relative to the case when no women smoked, as we computed earlier. We also obtain a 95% confidence interval of 34% to 58% for this estimate.

Example 4: Estimating the ATET

Intuitively, the ATET measures the effect of the treatment on an at-risk subpopulation. Sometimes the subpopulation that gets the treatment defines such an at-risk subpopulation. The ATET has the added benefit that it can be estimated under weaker conditions than the ATE; see Assumptions and tradeoffs in [CAUSAL] stteffects intro.
. stteffects ra (age exercise diet education) (smoke), atet
    Failure  _d:  fail
    Analysis time  _t:  atime

Iteration 0:  EE criterion = 1.006e-14
Iteration 1:  EE criterion = 2.970e-26

Survival treatment-effects estimation
Number of obs = 2,000
Estimator : regression adjustment
Outcome model : Weibull
Treatment model: none
Censoring model: none

|       _t | Robust Coefficient | std. err. | z  | P>|z| | [95% conf. interval] |
|---------|-------------------|-----------|----|------|----------------------|
| ATET    |                   |           |    |      |                      |
|   smoke |                   |           |    |      |                      |
| (Smoker |       -1.527476   | .2489203  | -6.14 | 0.000 | -2.015351 -1.039602 |
| vs      |                   |           |    |      |                      |
| Nonsmoker) |                |           |    |      |                      |
| POmean  |                   |           |    |      |                      |
| smoke   | 3.436937          | .2217808  | 15.50 | 0.000 | 3.002255 3.87162    |
| Nonsmoker) |                |           |    |      |                      |

When every woman in the subpopulation smoked, the average time to a second heart attack is estimated to be 1.53 years less than when no women in the subpopulation smoked. The estimated average time to a second heart attack when no women in the subpopulation smoked is 3.44 years.

\[\]

Example 5: Fixed or random censoring time

The time to censoring in survival-time data can be random or deterministic, although it must be independent of treatment assignment and the potential outcomes; see Kalbfleisch and Prentice (2002, chap. 3) for the standard case and see The correct adjustment for censoring assumption in Assumptions and tradeoffs in [CAUSAL] stteffects intro for the treatment-effects case.

The RA estimator and the likelihood-adjusted-censoring version of the inverse-probability-weighted RA estimator can accommodate a fixed time to censoring; see The correct adjustment for censoring assumption in [CAUSAL] stteffects intro. (The estimators that handle censoring by weighting cannot accommodate a fixed time to censoring because the weights are not well defined with a fixed time to censoring.)

We have fictional data on the time to rearrest among men aged 25–35 who were previously in prison for a felony conviction (rtime). The time to censoring is fixed in these data because individuals were followed for a maximum of five years.

Some of the young men chose to enter a vocational training program before release from prison; train is 1 for participants and 0 for nonparticipants. The dataset also contains fail (which is 1 if the observed time is a failure time and 0 if it is time to censoring), age at the time of the first arrest (age), an index of the parents’ socioeconomic level (parental), and the number of years behind in school at the time of the first arrest (edeficit).

We estimate the ATET because we wish to allow the gains from the training program to be related to an unobservable characteristic that affects who self-selects into the program; see Average treatment effect on the treated in [CAUSAL] stteffects intro.
We model the outcome as a function of age, parental, and edeficit.

```
use https://www.stata-press.com/data/r18/recid2, clear
(Time to rearrest (fictional))
.stteffects ra (age parental edeficit) (train), atet
   Failure _d: fail
   Analysis time _t: rtime
Iteration 0: EE criterion = 1.875e-23
Iteration 1: EE criterion = 4.785e-32
Survival treatment-effects estimation
Number of obs = 2,000
Estimator : regression adjustment
Outcome model : Weibull
Treatment model: none
Censoring model: none

     _t | Coefficient  std. err.     z  P>|z|      [95% conf. interval]
-----|-----------------+--------------------------------------
     ATET |                 |
     train |                 |
       (Student vs Nonstudent) | 2.440919   .4689057   5.21   0.000     1.52188   3.359957
     P0mean |                 |
     train |                 |
       Nonstudent | 2.062029   .1231492   16.74  0.000     1.820661   2.303397
```

When everyone who selected the training got the training, the average time to reaarrest is 2.44 years later than the average rearrest time if none of those who selected the training got the training. The average rearrest time if none of those who selected the training got the training is 2.06 years. In other words, the average time to rearrest increases from about 2.06 years to about 4.50 years for the subpopulation of young men who self-selected into the prerelease vocational training program.

### Stored results

**stteffects ra** stores the following in *e()*:

#### Scalars
- *e(N)*: number of observations
- *e(nj)*: number of observations for treatment level *j*
- *e(N_clust)*: number of clusters
- *e(k_eq)*: number of equations in *e(b)*
- *e(k_levels)*: number of levels in treatment variable
- *e(treated)*: level of treatment variable defined as treated
- *e(control)*: level of treatment variable defined as control
- *e(converged)*: 1 if converged, 0 otherwise

#### Macros
- *e(cmd)*: *stteffects*
- *e(cmdline)*: command as typed
- *e(dead)*: _d
- *e(depvar)*: _t
- *e(tvar)*: name of treatment variable
- *e(subcmd)*: ra
- *e(o_model)*: outcome model: weibull, exponential, gamma, or lognormal
- *e(stat)*: statistic estimated: ate, atet, or pomeans
Methods and formulas

The methods and formulas for the RA estimators implemented in \texttt{stteffects ra} are given in Methods and formulas of [CAUSAL] \texttt{stteffects ipwra}.

References


**Also see**

[CAUSAL] *stteffects postestimation* — Postestimation tools for stteffects

[CAUSAL] *stteffects intro* — Introduction to treatment effects for observational survival-time data

[ST] *streg* — Parametric survival models

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

[U] 20 *Estimation and postestimation commands*