Title

stata.com

stteffects ipw — Survival-time inverse-probability weighting

Description

`stteffects ipw` 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 with random time to censoring. Estimation is by inverse-probability weighting (IPW). IPW estimators use weighted averages of the observed outcome. The estimated weights correct for missing data on the potential outcomes and for censored survival times. `stteffects ipw` offers several choices for the functional forms of the treatment model and the time-to-censoring model. Binary and multivalued treatments are accommodated.

See [TE] **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 of binary `treat2` on `time` by IPW using a logistic model of `treat2` on `x` and `w` and using `x` and `w` in a Weibull model for the censoring time

```
stteffects ipw (treat2 x w) (x w)
```

As above, but estimate the ATET

```
stteffects ipw (treat2 x w) (x w), atet
```

ATE of `treat2` on `time` by IPW using a probit model of `treat2` on `x` and `w` and using `x` and `w` in a gamma model for the censoring time

```
stteffects ipw (treat2 x w, probit) (x w, gamma)
```

ATE for treatment levels 2 and 3 of three-valued treatment `treat3`

```
stteffects ipw (treat3 x w) (x w)
```

As above, and specify that `treat3` = 3 is the control level using the value label "MyControl" for 3

```
stteffects ipw (treat3 x w) (x w), control("MyControl")
```

Menu

Statistics > Treatment effects > Survival outcomes > Inverse-probability weighting (IPW)

Syntax

```
stteffects ipw (tvar tmvarlist [, tmoptions]) (cmvarlist [, cmoptions])
    [if] [in] [, stat options]
```

tvar must contain integer values representing the treatment levels.

tmvarlist specifies the variables that predict treatment assignment in the treatment model.

cmvarlist specifies the variables that predict censoring in the censoring model.

tmoptions	Description
Model	
`logit`	logistic treatment model; the default
`probit`	probit treatment model
`hetprobit(`*varlist*`)`	heteroskedastic probit treatment model
`noconstant`	suppress constant from treatment model

cmoptions	Description
Model	
`weibull`	Weibull; the default
`exponential`	exponential
`gamma`	two-parameter gamma
`lnormal`	lognormal
`ancillary(`*avarlist* [`, noconstant`]`)`	specify variables used to model ancillary parameter
`noconstant`	suppress constant from censoring model

stat	Description
Stat	
`ate`	estimate average treatment effect in population; the default
`atet`	estimate average treatment effect on the treated
`pomeans`	estimate potential-outcome means

options	Description
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `robust`, `cluster` *clustvar*, `bootstrap`, or `jackknife`
Reporting	
`level(#)`	set confidence level; default is `level(95)`
`aequations`	display auxiliary-equation results
`noshow`	do not show st setting information
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
`iterinit(#)`	specify starting-value iterations; seldom used
Advanced	
`pstolerance(#)`	set the tolerance for the overlap assumption
`osample(`*newvar*`)`	identify observations that violate the overlap assumption
`control(#` | *label*`)`	specify the level of *tvar* that is the control
`tlevel(#` | *label*`)`	specify the level of *tvar* that is the treatment
`coeflegend`	display legend instead of statistics

You must `stset` your data before using `stteffects`; see [ST] **stset**.

tmvarlist, *cmvarlist*, and *avarlist* may contain factor variables; see [U] **11.4.3 Factor variables**.

`bootstrap`, `by`, `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.

`coeflegend` does not appear in the dialog box.

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

Options

Model

`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

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.

SE/Robust

`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***.

Reporting

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

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

`noshow` prevents `stteffects ipw` 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`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(`*style*`)`, `cformat(%`*fmt*`)`, `pformat(%`*fmt*`)`, `sformat(%`*fmt*`)`, and `nolstretch`; see [R] **estimation options**.

Maximization

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.

Advanced

`pstolerance(#)` specifies the tolerance used to check the overlap assumption. The default value is `pstolerance(1e-5)`. `stteffects` will exit with an error if an observation has an estimated propensity score smaller than that specified by `pstolerance()`.

`osample(`*newvar*`)` specifies that indicator variable *newvar* be created to identify observations that violate the overlap assumption.

`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:

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

Remarks and examples

stata.com

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

IPW estimators use contrasts of weighted averages of observed outcomes to estimate treatment effects. The estimated weights correct for data that are missing because each subject is only observed after receiving one of the possible treatment levels and because some survival-time outcomes are censored.

The IPW estimators implemented in `stteffects ipw` use a three-step approach to estimating the ATE:

1. Estimate the parameters of a treatment-assignment model, and compute the component of the estimated weights that accounts for data missing because each subject is only observed after receiving one of the possible treatment levels.
2. Estimate the parameters of a time-to-censoring model, and compute the component of the estimated weights that accounts for data lost to censoring.
3. Use the estimated weights to compute weighted averages of the outcomes for each treatment level.

To estimate the ATET, we use different weights in step 2.

The time to censoring must be random to use `stteffects ipw` because the model in step 2 is not well defined if the time to censoring is fixed. See [TE] **stteffects intro** for more details. For information about estimators that accommodate a fixed time to censoring, see [TE] **stteffects ra** and [TE] **stteffects ipwra**.

Here we note only a few entry points to the vast literature on IPW estimators. Hirano, Imbens, and Ridder (2003), Imbens (2000, 2004), Imbens and Wooldridge (2009), Rosenbaum and Rubin (1983), Robins and Rotnitzky (2006), Wooldridge (2002, 2007), 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 the IPW estimators in particular.

Like `streg` and other survival-time commands, `stteffects ipw` uses the outcome variable and the failure indicator computed by, and optionally weights specified with, `stset`. `stteffects ipw` 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 [TE] **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 ipw` to estimate the ATE. We model treatment assignment using the default logit model with covariates on `age`, `exercise`, and `education`. We model the time to censoring using the default Weibull model with covariates on `age`, `exercise`, `diet`, and `education`.

```
. use http://www.stata-press.com/data/r15/sheart
(Time to second heart attack (fictional))
. stteffects ipw (smoke age exercise education) (age exercise diet education)
         failure _d:  fail
   analysis time _t:  atime
Iteration 0:   EE criterion =  2.042e-18
Iteration 1:   EE criterion =  5.191e-31
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
Censoring model: Weibull
```

_t	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
smoke (Smoker vs Nonsmoker)	-2.22226	.6307573	-3.52	0.000	-3.458522	-.9859983
POmean						
smoke Nonsmoker	4.235569	.5210937	8.13	0.000	3.214244	5.256894

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 2.22 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.

The ratio of the ATE to the control-level POM measures the importance of the effect. In this example, when every woman smoked, the average time to a second heart attack falls by an estimated 52% relative to the case when none of them smoked. See example 3 in [TE] **stteffects ra** for an example that uses `nlcom` to compute a point estimate and a confidence interval for this ratio.

◁

▷ Example 2: Different treatment and censoring models

Instead of a logit model for the treatment assignment, we could have used a probit or a heteroskedastic probit model. Instead of a Weibull model for the censoring time, we could have used an exponential, a gamma, or a lognormal model. For a quick comparison, we now estimate the ATE using a probit model for the treatment assignment and using a gamma model for the censoring time.

```
. stteffects ipw (smoke age exercise education, probit)
> (age exercise diet education, gamma)

        failure _d:  fail
  analysis time _t:  atime

Iteration 0:   EE criterion =  3.534e-15
Iteration 1:   EE criterion =  5.263e-27

Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: probit
Censoring model: gamma
```

_t	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
smoke (Smoker vs Nonsmoker)	-2.646808	.8368254	-3.16	0.002	-4.286956	-1.006661
POmean						
smoke Nonsmoker	4.702301	.7404567	6.35	0.000	3.251033	6.15357

The estimated ATE of −2.65 and control-level POM of 4.70 are similar to the values of −2.22 and 4.24 reported in example 1.

◁

▷ Example 3: 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 trade-offs* under *Remarks and examples* in [TE] **stteffects intro**.

```
. stteffects ipw (smoke age exercise education) (age exercise diet education),
> atet

         failure _d:  fail
   analysis time _t:  atime

Iteration 0:   EE criterion =  2.042e-18
Iteration 1:   EE criterion =  1.248e-32

Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
Censoring model: Weibull
```

_t	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATET						
smoke (Smoker vs Nonsmoker)	-1.846136	.5076872	-3.64	0.000	-2.841185	-.8510877
POmean						
smoke Nonsmoker	3.543788	.474395	7.47	0.000	2.613991	4.473585

When every woman in the subpopulation smoked, the average time to a second heart attack is estimated to be 1.85 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 3.54 years. ◁

Stored results

`stteffects ipw` stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(n`*j*`)`	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)`	`ipw`
`e(tmodel)`	treatment model: `logit`, `probit`, or `hetprobit`
`e(cmodel)`	censoring model: `weibull`, `exponential`, `gamma`, or `lognormal`
`e(stat)`	statistic estimated: `ate`, `atet`, or `pomeans`
`e(wtype)`	weight type
`e(wexp)`	weight expression
`e(title)`	title in estimation output
`e(clustvar)`	name of cluster variable
`e(tlevels)`	levels of treatment variable

e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. Err.
e(properties)	b V
e(estat_cmd)	program used to implement estat
e(predict)	program used to implement predict
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(V)	variance–covariance matrix of the estimators
Functions	
e(sample)	marks estimation sample

Methods and formulas

The methods and formulas for the IPW estimators implemented in stteffects ipw are given in *Methods and formulas* of [TE] **stteffects ipwra**.

References

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bai, X., A. A. Tsiatis, and S. M. O'Brien. 2013. Doubly robust estimators of treatment-specific survival distributions in observational studies with stratified sampling. *Biometrics* 69: 830–839.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cattaneo, M. D. 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* 155: 138–154.

Cattaneo, M. D., D. M. Drukker, and A. D. Holland. 2013. Estimation of multivalued treatment effects under conditional independence. *Stata Journal* 13: 407–450.

Guo, S., and M. W. Fraser. 2015. *Propensity Score Analysis: Statistical Methods and Applications*. 2nd ed. Thousand Oaks, CA: Sage.

Hirano, K., G. W. Imbens, and G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71: 1161–1189.

Imbens, G. W. 2000. The role of the propensity score in estimating dose–response functions. *Biometrika* 87: 706–710.

——. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics* 86: 4–29.

Imbens, G. W., and J. M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47: 5–86.

Robins, J. M., and A. Rotnitzky. 2006. Inverse probability weighting in survival analysis. In *Survival and Event History Analysis*, ed. N. Keiding and P. K. Andersen, 266–271. Chichester, UK: Wiley.

Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701.

Tsiatis, A. A. 2006. *Semiparametric Theory and Missing Data*. New York: Springer.

Vittinghoff, E., D. V. Glidden, S. C. Shiboski, and C. E. McCulloch. 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. New York: Springer.

Wooldridge, J. M. 2002. Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal* 1: 117–139.

——. 2007. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* 141: 1281–1301.

——. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Also see

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

[TE] **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**