teffects ipw — Inverse-probability weighting

Description

teffects ipw estimates the average treatment effect (ATE), the average treatment effect on the treated (ATET), and the potential-outcome means (POMs) from observational data by inverse-probability weighting (IPW). IPW estimators use estimated probability weights to correct for missing data on the potential outcomes. teffects ipw accepts a continuous, binary, count, fractional, or nonnegative outcome and allows a multivalued treatment.

See [TE] teffects intro or [TE] teffects intro advanced for more information about estimating treatment effects from observational data.

Quick start

ATE of binary treat2 on y by IPW using a logistic model of treat2 on x and w

teffects ipw (y) (treat2 x w)

As above, but estimate ATET

teffects ipw (y) (treat2 x w), atet

As above, but estimate potential-outcome means

teffects ipw (y) (treat2 x w), pomeans

ATE of treat2 on y using heteroskedastic probit for treat2 as a function of x and w

teffects ipw (y) (treat2 x w, hetprobit(x w))

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

teffects ipw (y) (treat3 x w)

As above, and specify that treat3 = 3 is the control level

teffects ipw (y) (treat3 x w), control(3)

Same as above, specified using the label “MyControl” corresponding to treat3 = 3

teffects ipw (y) (treat3 x w), control(MyControl)

Menu

Statistics > Treatment effects > Continuous outcomes > Inverse-probability weighting (IPW)
Statistics > Treatment effects > Binary outcomes > Inverse-probability weighting (IPW)
Statistics > Treatment effects > Count outcomes > Inverse-probability weighting (IPW)
Statistics > Treatment effects > Fractional outcomes > Inverse-probability weighting (IPW)
Statistics > Treatment effects > Nonnegative outcomes > Inverse-probability weighting (IPW)


## Syntax

```
teffects ipw (ovar) (tvar tmvarlist [, tmodel noconstant]) [if] [in] [weight] [, stat options]
```

**ovar** is a binary, count, continuous, fractional, or nonnegative outcome of interest.

**tvar** must contain integer values representing the treatment levels.

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

### Model

<table>
<thead>
<tr>
<th><strong>tmodel</strong></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>logit</td>
<td>logistic treatment model; the default</td>
</tr>
<tr>
<td>probit</td>
<td>probit treatment model</td>
</tr>
<tr>
<td>hetprobit(varlist)</td>
<td>heteroskedastic probit treatment model</td>
</tr>
</tbody>
</table>

**tmodel** specifies the model for the treatment variable.

For multivalued treatments, only **logit** is available and multinomial logit is used.

### Stat

<table>
<thead>
<tr>
<th><strong>stat</strong></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ate</td>
<td>estimate average treatment effect in population; the default</td>
</tr>
<tr>
<td>atet</td>
<td>estimate average treatment effect on the treated</td>
</tr>
<tr>
<td>pomeans</td>
<td>estimate potential-outcome means</td>
</tr>
</tbody>
</table>

### Options

<table>
<thead>
<tr>
<th><strong>options</strong></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>vce(vcetype)</td>
<td>vcetype may be <strong>robust</strong>, <strong>cluster clustvar</strong>, <strong>bootstrap</strong>, or <strong>jackknife</strong></td>
</tr>
</tbody>
</table>

**Reporting**

<table>
<thead>
<tr>
<th><strong>level(#)</strong></th>
<th>set confidence level; default is level(95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aequations</td>
<td>display auxiliary-equation results</td>
</tr>
<tr>
<td>display_options</td>
<td>control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
</tbody>
</table>

**Maximization**

| **maximize_options** | control the maximization process; seldom used |

**Advanced**

<table>
<thead>
<tr>
<th><strong>pстolerance(#)</strong></th>
<th>set tolerance for overlap assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>osample(newvar)</td>
<td>newvar identifies observations that violate the overlap assumption</td>
</tr>
<tr>
<td>control(#</td>
<td>label)</td>
</tr>
<tr>
<td>tlevel(#</td>
<td>label)</td>
</tr>
<tr>
<td>coeflegend</td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>
tmvarlist 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 are allowed; see [U] 11.1.6 weight.

coefflegend does not appear in the dialog box.

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

**Options**

---

**Model**

-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.

eaequations 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.

display_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, 
allbaselevels, nolabel, 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

---

**Advanced**

-pstolerance(#) specifies the tolerance used to check the overlap assumption. The default value is pstolerance(1e-5). teffects 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 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 teffects ipw but is not shown in the dialog box: coeflegend; see [R] Estimation options.

Remarks and examples

Remarks are presented under the following headings:

  Overview
  Video example

Overview

IPW estimators use estimated probability weights to correct for the missing-data problem arising from the fact that each subject is observed in only one of the potential outcomes. IPW estimators use a two-step approach to estimating treatment effects:

1. They estimate the parameters of the treatment model and compute the estimated inverse-probability weights.
2. They use the estimated inverse-probability weights to compute weighted averages of the outcomes for each treatment level. The contrasts of these weighted averages provide the estimates of the ATEs. Using this weighting scheme corrects for the missing potential outcomes.

These steps produce consistent estimates of the effect parameters because the treatment is assumed to be independent of the potential outcomes after conditioning on the covariates. The overlap assumption ensures that predicted inverse-probability weights do not get too large. In fact, teffects ipw uses an estimation technique that implements both steps at once so that we do not need to correct the standard errors in the second step to reflect the uncertainty associated with the predicted treatment probabilities.

We will illustrate the use of teffects ipw by using data from a study of the effect of a mother’s smoking status during pregnancy (mbsmoke) on infant birthweight (bweight) as reported by Cattaneo (2010). This dataset also contains information about each mother’s age (mage), education level (medu), marital status (mmarried), whether the first prenatal exam occurred in the first trimester (prenatal1), and whether this baby was the mother’s first birth (fbaby).

Example 1: Estimating the ATE

We begin by using teffects ipw to estimate the average treatment effect of smoking on birthweight. We will use a probit model to predict treatment status, using prenatal1, mmarried, mage, the square of mage, and fbaby as explanatory variables:
. use https://www.stata-press.com/data/r16/cattaneo2
. teffects ipw (bweight) (mbsmoke mmarried c.mage##c.mage fbaby medu, probit)
Iteration 0:   EE criterion =  4.621e-21
Iteration 1:   EE criterion =  7.358e-26
Treatment-effects estimation                     Number of obs   =    4,642
Estimator             : inverse-probability weights
Outcome model        : weighted mean
Treatment model: probit

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td><strong>ATE</strong></td>
<td></td>
</tr>
<tr>
<td>mbsmoke (smoker vs</td>
<td>-230.6886</td>
</tr>
<tr>
<td>nonsmoker)</td>
<td></td>
</tr>
<tr>
<td><strong>POmean</strong></td>
<td></td>
</tr>
<tr>
<td>mbsmoke nonsmoker</td>
<td>3403.463</td>
</tr>
</tbody>
</table>

The average birthweight if all mothers were to smoke would be 231 grams less than the average of 3,403 grams that would occur if none of the mothers had smoked.

Sometimes, we are mainly concerned about those subjects that did in fact receive treatment, and we want to know how much the outcome changes as a result of treatment for that subpopulation. The ATET provides us with the answer. Moreover, the ATET can be estimated using weaker assumptions than are required to estimate the ATE; see [TE] teffects intro advanced.

> Example 2: Estimating the ATET

. teffects ipw (bweight) (mbsmoke mmarried c.mage##c.mage fbaby medu, probit),
> atet
Iteration 0:   EE criterion =  4.636e-21
Iteration 1:   EE criterion =  6.467e-27
Treatment-effects estimation                     Number of obs   =    4,642
Estimator             : inverse-probability weights
Outcome model        : weighted mean
Treatment model: probit

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td><strong>ATET</strong></td>
<td></td>
</tr>
<tr>
<td>mbsmoke (smoker vs</td>
<td>-225.1773</td>
</tr>
<tr>
<td>nonsmoker)</td>
<td></td>
</tr>
<tr>
<td><strong>POmean</strong></td>
<td></td>
</tr>
<tr>
<td>mbsmoke nonsmoker</td>
<td>3362.837</td>
</tr>
</tbody>
</table>
The average birthweight is 225 grams less when all the mothers who smoke do so than the average of 3,363 grams that would have occurred if none of these mothers had smoked.

We often express statistics as percentages to alleviate scaling issues and aid interpretation. In the present context, we may wish to express an ATE as a percentage of the untreated POM to gain a more intuitive measure of the effect of treatment.

### Example 3: Reporting the ATE as a percentage

Here we use the same model as in example 1, but we report the ATE as a percentage of the mean birthweight that would occur if no mothers smoke. First, we use `teffects ipw` to fit the model. We use the `coeflegend` option so that `teffects ipw` reports the names of the parameters. Then we use `nlcom` to obtain the statistic we want along with its delta-method-based standard error. We type

```
. teffects ipw (bweight) (mbsmoke mmarried c.mage##c.mage fbaby medu, probit), > coeflegend
```

```
Iteration 0: EE criterion = 4.621e-21
Iteration 1: EE criterion = 7.358e-26
Treatment-effects estimation Number of obs = 4,642
Estimator : inverse-probability weights
Outcome model : weighted mean
Treatment model: probit

<table>
<thead>
<tr>
<th>bweight</th>
<th>Coef. Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATE</td>
<td></td>
</tr>
<tr>
<td>mbsmoke (smoker vs nonsmoker)</td>
<td>-230.6886 _b[ATE:r1vs0.mbsmoke]</td>
</tr>
<tr>
<td>POmean mbsmoke nonsmoker</td>
<td>3403.463 _b[POmean:0.mbsmoke]</td>
</tr>
</tbody>
</table>
```

```
.nlcom _b[ATE:r1vs0.mbsmoke] / _b[POmean:0.mbsmoke]
  _nl_1: _b[ATE:r1vs0.mbsmoke] / _b[POmean:0.mbsmoke]
```

```
| bweight | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------|-------|-----------|-------|------|---------------------|
| _nl_1  | -.0677806 | .0075169 | -9.02 | 0.000 | -0.0825133 -0.0530478 |
```

The average birthweight falls by an estimated 6.8% when every mother smokes relative to the case when no mothers smoke. We also obtain a 95% confidence interval of a 5.3% to 8.3% reduction.

### Video example

**Treatment effects: Inverse-probability weighting**
Stored results

teffects 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): teffects
- e(cmdline): command as typed
- e(depvar): name of outcome variable
- e(tvar): name of treatment variable
- e(subcmd): ipw
- e(tmodel): logit, probit, or hetprobit
- e(stat): statistic estimated, ate, atet, or pomeans
- e(vtype): 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

teffects ipw implements a smooth treatment-effects estimator. All smooth treatment-effects estimators are documented in Methods and formulas of [TE] teffects aipw.

References


Also see

[TE] `teffects postestimation` — Postestimation tools for `teffects`

[TE] `teffects` — Treatment-effects estimation for observational data

[U] 20 Estimation and postestimation commands