

teffects intro — Introduction to treatment effects for observational data

[Description](#)[Remarks and examples](#)[Reference](#)[Also see](#)

Description

This entry provides a nontechnical introduction to treatment-effects estimators and the `teffects` command in Stata. Advanced users may want to instead read [\[TE\] `teffects intro advanced`](#) or skip to the individual commands' entries.

The `teffects` command estimates potential-outcome means (POMs), average treatment effects (ATEs), and average treatment effects among treated subjects (ATETs) using observational data.

Treatment effects can be estimated using regression adjustment (RA), inverse-probability weights (IPW), and “doubly robust” methods, including inverse-probability-weighted regression adjustment (IPWRA) and augmented inverse-probability weights (AIPW), and via matching on the propensity score or nearest neighbors.

The outcome models can be continuous, binary, count, or nonnegative. Continuous outcomes can be modeled using linear regression; binary outcomes can be modeled using logit, probit, or heteroskedastic probit regression; and count and nonnegative outcomes can be modeled using Poisson regression. The treatment model can be binary or multinomial. Binary treatments can be modeled using logit, probit, or heteroskedastic probit regression, while multinomial outcomes are modeled using multinomial logit regression.

Remarks and examples

stata.com

This entry presents a nontechnical overview of treatment-effects estimators for those who are new to the subject of treatment-effects estimation or are at least new to Stata's facilities for estimating treatment effects. More advanced users may want to instead read [\[TE\] `teffects intro advanced`](#) or skip to the individual commands' entries.

Remarks are presented under the following headings:

[Introduction](#)[Defining treatment effects](#)[Estimating treatment effects](#)[A quick tour of the estimators](#)[Video example](#)

Introduction

Suppose we have observed a sample of subjects, some of whom received a treatment and the rest of whom did not. As the name suggests, in most applications, the “subjects” are indeed people. A “treatment” could indeed be a medical treatment such as a new drug regimen or surgical procedure. In social science applications, a treatment could be participation in a job-training program or inclusion in a classroom or school in which a new pedagogical method is being used. However, not all applications use individuals as the subjects. For example, a policy analyst might be interested in examining the impact of an experimental program in which a national agency held a lottery to award only some local governments the resources needed to implement the program. Here the subjects are the local governments, and treatment refers to whether a local government received the resources needed to implement the program.

We would like to know if a treatment has an effect on an outcome Y . The outcome could be the cholesterol level of a patient taking either an existing statin or a new experimental drug, or the outcome could be the wage offered to a person who either did or did not participate in a job-training program. In an ideal world, we would observe Y when a subject is treated (which we denote as Y_1), and we would observe Y when the same subject is not treated (which we denote as Y_0). We would be careful to make both observations under identical conditions so that the only difference is the presence or absence of the treatment. We could then average the difference between Y_1 and Y_0 across all the subjects in our dataset to obtain a measure of the average impact of the treatment.

Unfortunately, this ideal experiment is almost never available in observational data because it is not possible to observe a specific subject having received the treatment and having not received the treatment. When the outcome is the birthweight of a specific baby and the treatment is the mother smoking while pregnant, it is impossible to observe the baby's birthweight under both treatments of the mother smoking and the mother not smoking.

A classic solution to this problem is to randomize the treatment. High costs or ethical issues rule out this solution in many observational datasets. For example, we could not ask a random selection of pregnant women to smoke.

The defining characteristic of observational data is that treatment status is not randomized. Moreover, that implies that the outcome and treatment are not necessarily independent. The goal of the estimators implemented by `teffects` is to utilize covariates to make treatment and outcome independent once we condition on those covariates.

The treatment-effect estimators implemented by `teffects` allow us to estimate the efficacy of treatments using observational data. The rest of this entry discusses these treatment-effect estimators at an introductory level. For a more technical introduction, see [\[TE\] `teffects` intro advanced](#).

Defining treatment effects

We introduce treatment effects more formally by using the potential-outcomes framework, which is also known as the counterfactual framework. What is a potential outcome? Consider a subject that did not receive treatment so that we observe Y_0 . What would Y_1 be for that same subject if it were exposed to treatment? We call Y_1 the potential outcome or counterfactual for that subject. For a subject that did receive treatment, we observe Y_1 , so Y_0 would be the counterfactual outcome for that subject. We can view this as a missing-data problem, and treatment-effect methods can account for that problem.

Treatment-effect estimators allow us to estimate three parameters. The potential-outcome means (POMs) are the means of Y_1 and Y_0 in the population. The average treatment effect (ATE) is the mean of the difference ($Y_1 - Y_0$). Finally, the average treatment effect on the treated (ATET) is the mean of the difference ($Y_1 - Y_0$) among the subjects that actually receive the treatment.

To develop our intuition, suppose we have observed a sample of patients, some of whom received a medication to reduce their blood pressure. [Figure 1](#) plots each of our patient's systolic blood pressures as a function of weight. We use the color green to indicate patients who did not receive the drug and blue to indicate patients who did receive the drug.

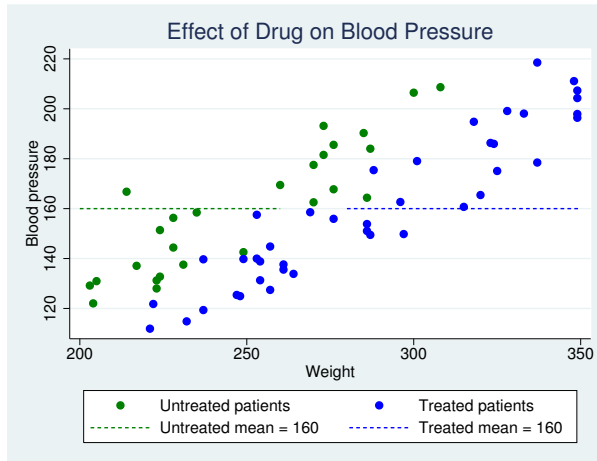


Figure 1

A remarkable feature of our data is that the average blood pressure of patients not taking the drug is 160, and the average blood pressure of patients taking the drug is also 160. Can we therefore conclude that taking the drug has no impact on blood pressure? The answer is no.

Because this is observational data, we could not randomly assign who would receive the drug and who would not. As a result, treatment status could be related to covariates that also affect blood pressure. Heavier patients were more likely to be prescribed the medication, and blood pressure is correlated with weight. The difference in sample means does not estimate the true average treatment effect, because blood pressure depends on weight and weight is correlated with the treatment.

Suppose that we did in fact observe both potential outcomes for all patients. In [figure 2](#), we continue to use solid dots for our observed data points, and we introduce hollow dots to represent the counterfactual outcomes. That is, the green hollow dots represent the blood pressures we would measure if only our treated patients had not taken the drug, and the blue hollow dots represent the blood pressures we would measure if only our untreated patients had taken the drug. The green and blue dashed lines represent the untreated and treated POMs, respectively. That is, the green line represents the mean of all the green dots, and the blue line represents the mean of all the blue dots.

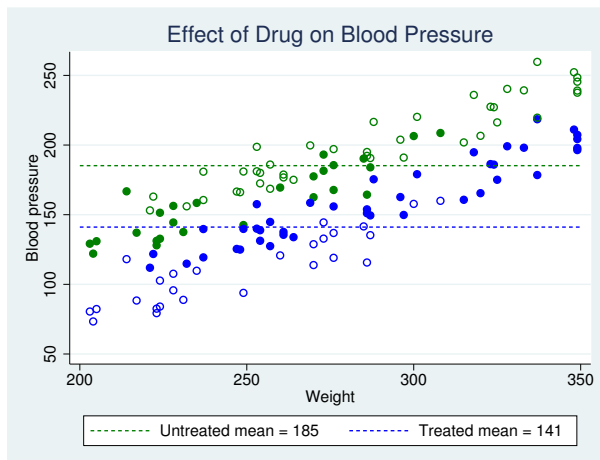


Figure 2

If we did have the data represented by the hollow dots, then we could say that the average treatment effect is the difference between the mean of all the green dots and the mean of all the blue dots. In this ideal scenario, there are no missing data on the other potential outcome, and we have all the data we need to use the difference in means to estimate the ATE.

Looking at figure 2, we can see why a difference in means using only the solid dots does not estimate the ATE. Using only the solid green dots underestimates the average blood pressure for untreated individuals, and using only the solid blue dots overestimates the average blood pressure for treated individuals.

Estimating an ATE is essentially a missing-data problem. When covariates that affect the potential outcomes are related to treatment, we cannot use a difference in sample means, because the missing data are informative.

The treatment-effect estimators implemented in `teffects` allow for covariates like weight to be related to the potential outcomes and the treatment. Essentially, the estimators implemented by `teffects` utilize covariates to fill in the hollow circles or otherwise account for how the missing data depend on covariates that affect the potential outcomes.

Estimating treatment effects

We cannot estimate the ATE by simply taking the difference between the sample means for the treated and untreated subjects, because there are covariates that are related to the potential outcomes and the treatment. The estimators implemented by `teffects` require us to specify enough of these covariates so that after we condition on these covariates, any remaining influences on the treatment are not related to the potential outcomes. `teffects` implements several different estimators to accomplish this, including regression adjustment (RA), inverse-probability weighting (IPW), “doubly robust” methods that combine elements of RA and IPW, and matching methods. Here we introduce the methods by using intuition and simple examples.

See [TE] [teffects intro advanced](#) for a more technical introduction, and see the individual commands’ entries for estimator-specific details.

Regression adjustment

The RA method extends the idea of using sample means to estimate treatment effects by using a regression model to predict potential outcomes adjusted for covariates. In the examples here, we use linear regression, but the `teffects ra` command provides you with the flexibility to use logistic, probit, and heteroskedastic probit regression models for binary outcomes as well as Poisson regression for nonnegative outcomes; see [TE] [teffects ra](#) for more information.

`bweightex.dta` is a hypothetical dataset based on [Cattaneo \(2010\)](#) that we have created to illustrate treatment-effects estimators using graphs. The subjects in this dataset are women who were pregnant, some of whom smoked during the pregnancy. The outcome variable is the birthweight of the baby, and we want to know whether smoking during pregnancy affects the birthweight. The dataset also contains other demographic variables that we will use later.

Figure 3 illustrates the relationship between birthweight and smoking status as a function of the mother's age:

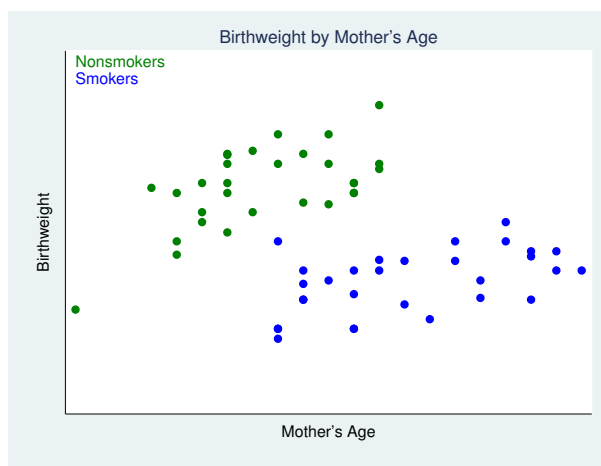


Figure 3

We see that smokers tend to be older than nonsmokers and that birthweight depends on smoking. Therefore, the difference between the sample means of birthweights of babies born to smokers and nonsmokers will not estimate the true average treatment effect.

We also still have the same problem as in the previous section: we do not observe the counterfactual birthweights of babies. Suppose, however, that we did. In [figure 4](#), we use solid points to represent observed birthweights and the colors green to represent nonsmokers and blue to represent smokers. The hollow points represent the counterfactual birthweights. The hollow blue points represent the birthweights of babies that we would observe if only our young nonsmoking mothers had instead smoked during their pregnancies. Similarly, the hollow green points represent the birthweights of babies that we would observe if only our older smoking mothers had instead not smoked during their pregnancies.

Figure 4 suggests a way to estimate the potential outcomes for each mother:

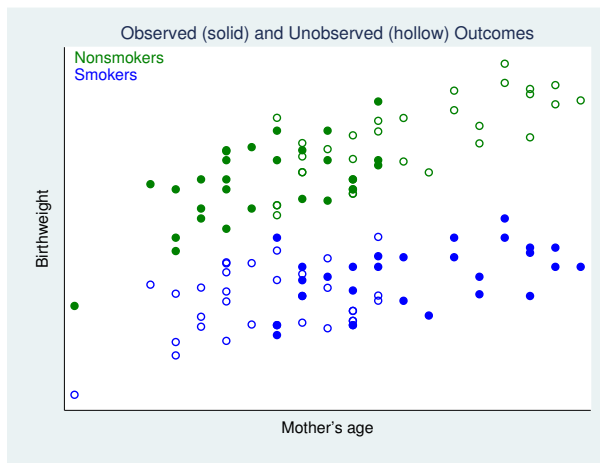


Figure 4

We could fit a linear regression of birthweight on mother's age by using the observed birthweights for nonsmokers, and we could do likewise for smokers. The following graph includes these two regression lines:

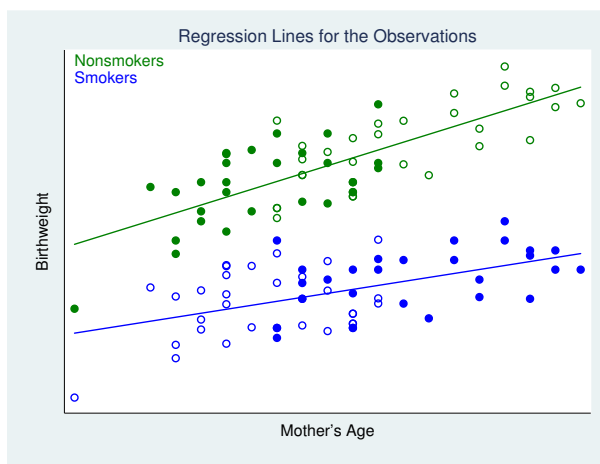


Figure 5

Figure 5 illustrates the principle behind the RA method. We use the green regression line to predict each baby's birthweight assuming the mother did not smoke, and we use the blue regression line to predict each baby's birthweight assuming the mother did smoke. The treatment effect of smoking for a mother of a particular age is the vertical difference between the green and blue regression lines.

The three parameters we mentioned in the introduction are now easy to estimate. For each mother, we obtain two values, say, bw_0 and bw_1 , representing our predictions of her baby's birthweight assuming the mother did not or did smoke, respectively. The means of these variables represent the untreated and treated POMs. The ATE is the sample mean of the difference $(bw_1 - bw_0)$, and the ATET is the sample mean of that difference computed using only the mothers who in fact did smoke during pregnancy.

Adding the circles highlights the fact that the average age is higher for smokers than for nonsmokers. Even though the blue and green lines have different slopes, if the average age was the same for smokers and nonsmokers, a difference in the sample means of birthweights could still estimate the true ATE.

Figure 5 lets us address one more issue. Users who are versed in regression analysis may be inclined to estimate the effect of smoking using a regression model for birthweight as a function of smoking and the mother's age. We clearly see in figure 5 that regression lines for smokers and nonsmokers have different slopes—the effect of age on birthweight is not the same for smokers and nonsmokers. In regression analysis, we would therefore include an interaction term between smoking and age. The RA method fits separate regression lines for smokers and nonsmokers, which also handles these differential effects of age on smoking.

Inverse-probability weighting

As we remarked in our discussion of the RA method, we cannot simply use the sample mean birthweights of babies born to smokers and nonsmokers to estimate the effect of smoking. If we did that, we would conflate the negative effect of smoking with the positive effect of age and the positive relationship between age and smoking. IPW is a treatment-effects estimator that uses weighted means rather than simple unweighted means to disentangle the effects of treatment and other confounders like age.

The concept underlying IPW can be gleaned from figure 2, where, as you will recall, the hollow points represent counterfactual outcomes. As we demonstrated in *Defining treatment effects*, we could estimate the average treatment effect if we knew the means of all the nonsmoking outcomes and the means of all the smoking outcomes. In the context of figure 4, we need the mean of all the green points, both solid and hollow, and the mean of all the blue points, both solid and hollow. Then the ATE would be the difference between those two means. IPW estimators view the hollow circles as missing data and correct the estimators for treated and not treated sample means for the missing data.

If we tried to calculate the mean nonsmoking birthweight based on just the solid green points, that mean would be biased downward because we are ignoring all the hollow green points, which correspond to higher birthweights. In IPW, we therefore apply more weight to the solid green points corresponding to older mothers and less weight to those corresponding to younger mothers. Using this weighting scheme will pull up the estimated mean birthweight of babies born to nonsmoking mothers to better approximate the true mean of all nonsmoking outcomes. The method for the mean smoking birthweight is virtually the same: we need to apply more weight to the younger smoking mothers than to the older smoking mothers to better approximate the true mean of all smoking outcomes.

Where do these weights for the weighted means come from? As the name implies, IPW uses the reciprocal (inverse) of the probability of being in the observed treatment group. These probabilities are obtained by fitting a model of treatment status on whatever characteristics we have on each subject. In our exposition of the RA method, we focused solely on the mother's age and smoking status as determinants of each baby's birthweight, mainly so that we could provide a graphical analysis. In the same vein, we could fit a logistic regression model of smoking status using only age to predict the probability that each mother smoked during pregnancy. However, the utility of the IPW method depends on how accurately the treatment model predicts the probability of treatment. In our dataset, we have other variables such as alcohol consumption, marital status, and the education level of the baby's father that are likely to help explain whether the mother smoked during pregnancy. As an added bonus, we can still provide a graphical exposition of IPW even with these additional explanatory variables.

We first fit a logistic model of the mother's smoking status, `mbsmoke`, as a function of marital status (`mmarried`), alcohol consumption (`alcohol`), the education level of the father (`fedu`), and the mother's age (`mage`):

```
. use http://www.stata-press.com/data/r13/bweightex
(Hypothetical birthweight data)
. logistic mbsmoke mmarrried alcohol mage fedu
Logistic regression
```

	Number of obs	=	60
	LR chi2(4)	=	46.50
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.5590

```
Log likelihood = -18.339432
```

mbsmoke	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
<code>mmarrried</code>	.0785086	.0909212	-2.20	0.028	.0081122	.7597976
<code>alcohol</code>	18.81727	27.98003	1.97	0.048	1.020649	346.9259
<code>mage</code>	2.147569	.459327	3.57	0.000	1.41218	3.265909
<code>fedu</code>	.8189843	.1157528	-1.41	0.158	.6208252	1.080393
<code>_cons</code>	4.46e-07	2.12e-06	-3.07	0.002	3.96e-11	.0050329

Next we compute the inverse-probability weights, which we will store in a variable called `ps`. In the IPW method, for subjects who did receive treatment, the weight is equal to the reciprocal of the predicted probability of treatment. For subjects who did not receive treatment, the weight is equal to the reciprocal of the predicted probability of not receiving treatment; the probability of not receiving treatment is just one minus the probability of receiving treatment:

```
. predict ps
(option pr assumed; Pr(mbsmoke))
. replace ps = 1/ps if mbsmoke==1
(30 real changes made)
. replace ps = 1/(1-ps) if mbsmoke==0
(30 real changes made)
```

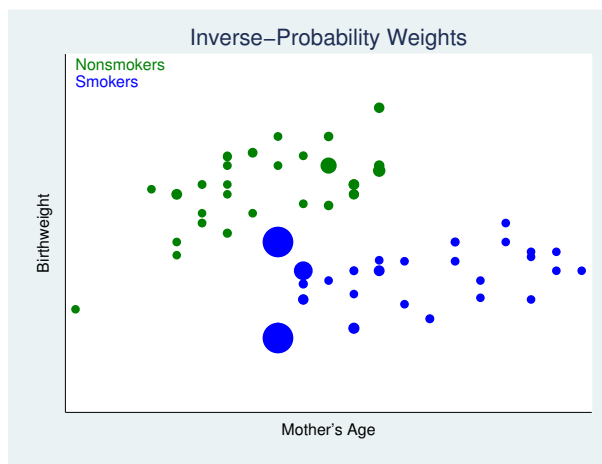


Figure 6

Figure 6 replicates [figure 3](#) with one twist. Rather than making all the points the same size, we have made the size of the points proportional to the IPW variable `ps`. Notice that the largest blue points correspond to the youngest smoking mothers in our sample, so they will receive the most weight when we compute the weighted mean birthweight of babies born to smoking mothers, just

as we explained we wanted to do. Similarly, the green points corresponding to older nonsmoking mothers are larger, representing larger weights.

There is a caveat to the IPW estimator. We remarked that in our sample, we had very few young mothers who smoked. When we fit our logistic or probit model and obtain the predicted probabilities, it is possible that some of the predictions are close to zero. The IPW is the reciprocal of that probability, and as the probability goes to zero, the weight becomes arbitrarily large. In those cases, the IPW can become unstable. This phenomenon is related to the concept of overlap, which means that every subject must have a positive probability of obtaining treatment. See [TE] [teffects overlap](#) and [TE] [teffects intro advanced](#) for more information about overlap.

Another limitation of the IPW estimator is that because we are using weighted means to estimate the POMs and ATE, unlike the RA estimator, we cannot obtain subject-level predictions of the treatment effects or potential outcomes. We do not have the two regression lines that we can use to predict outcomes for each subject.

Doubly robust combinations of RA and IPW

You may have noticed a clear distinction between the RA and IPW estimators. In the case of RA, we built linear regression models to predict the outcomes (birthweights) of each subject but said nothing about how treatment (smoking) arises. In the case of IPW, we built a logistic regression model to predict treatment status but did not build a formal model of the outcome. Doubly robust estimators combine the outcome modeling strategy of RA and the treatment modeling strategy of IPW. These estimators have a remarkable property: although they require us to build two models, we only need to specify one of the two models correctly. If we misspecify the treatment model but correctly specify the outcome model, we still obtain correct estimates of the treatment effect. If we correctly specify the treatment model but misspecify the outcome model, we again will obtain correct estimates of the treatment effect.

Stata's `teffects` command implements two doubly robust estimators, the augmented inverse-probability-weighted (AIPW) estimator and the inverse-probability-weighted regression-adjustment (IPWRA) estimator. These estimators combine elements of RA and IPW to be more robust to misspecification.

The AIPW estimator is an IPW estimator that includes an augmentation term that corrects the estimator when the treatment model is misspecified. When the treatment model is correctly specified, the augmentation term vanishes as the sample size becomes large. Like the IPW, the AIPW does not perform well when the predicted treatment probabilities are too close to zero or one.

The IPWRA estimator is an RA estimator that uses estimated inverse-probability weights to correct the estimator when the regression function is misspecified. When the regression function is correctly specified, the weights do not affect the consistency of the estimator.

Matching

Matching estimators are based on the idea of comparing the outcomes of subjects that are as similar as possible with the sole exception of their treatment status. In our birthweight and smoking example, we could select a mother who smokes and select a mother of the same age who does not smoke and compare the birthweights of their infants. The data of each mother serve as the potential outcome for the other mother.

For a single covariate such as age, identifying a pair of comparable mothers is not difficult. If we have a second covariate that is categorical, such as race, we might still be able to identify pairs of mothers who are the same age and of the same race assuming our dataset is large enough. However,

once we consider covariates that are measured on continuous scales or allow for more than a few discrete ones, then finding identical matches is a challenge. The solution is to use what is called a similarity measure, which is a statistic that measures how “close” two observations are. **teffects** offers two methods to find comparable observations based on similarity measures: nearest-neighbor matching and propensity-score matching.

Nearest-neighbor matching (NNM) is accomplished by calculating the “distance” between pairs of observations with regard to a set of covariates and then “matching” each subject to comparable observations that are closest to it. For example, suppose we have a variable that records each subject’s annual income to the penny. Say one subject who received treatment had an income of \$69,234.21. The likelihood that our dataset has an untreated subject who also earned \$69,234.21 is nil. However, we can determine the difference between each untreated subject’s income and our treated subject’s income, then match our treated subject with the untreated subjects whose income differences are smallest. Measuring the distance between subjects when we have multiple covariates is no challenge. By default, **teffects** uses what is known as the Mahalanobis distance, which is really nothing more than the Pythagorean theorem adapted to handle the fact that covariates may be correlated and measured on different scales.

NNM does not use a formal model for either the outcome or the treatment status, but this flexibility comes at a price. When matching on more than one continuous covariate, the NNM estimator must be augmented with a bias-correction term. **teffects nmatch** uses a linear function of the covariates specified in the `biasadj()` option to remove the large-sample bias.

Propensity-score matching (PSM) is an alternative to NNM. PSM matches on the estimated predicted probabilities of treatment, known as the propensity scores. PSM does not require bias correction, because it uses a model for the treatment. If the treatment model is reasonably well specified, PSM will perform at least as well as NNM; see [\[TE\] teffects intro advanced](#).

Caveats and assumptions

To use the estimators implemented in **teffects**, we must make several assumptions about the process that generated our data. Different estimators and statistics may require slightly more or slightly less restrictive assumptions and may exhibit varying degrees of robustness to departures from these assumptions, but in general, all the estimators require some form of the following three assumptions.

The independent and identically distributed (i.i.d.) sampling assumption ensures that the outcome and treatment status of each individual are unrelated to the outcome and treatment status of all the other individuals in the population. Correlated data arising from hierarchical or longitudinal study designs do not meet this assumption.

The conditional-independence (CI) assumption means once we control for all observable variables, the potential outcomes are independent of treatment assignment. The easiest way to understand the CI assumption is to understand when it is violated. In our birthweight example, suppose mothers who did not smoke were more health conscious and consumed better prenatal diets than those who did smoke. Unless we explicitly controlled for health awareness or diet, our model would violate the CI assumption: the mother’s decision to smoke or not smoke would not be independent of the baby’s birthweight. If we did not control for health awareness, we would overstate the negative impact of smoking on birthweight. Babies born to mothers who smoke weigh less than babies born to nonsmoking mothers not just because of the effects of cigarettes but also because of poorer prenatal diets.

In a study examining the effect of a job-training program, the CI assumption requires that there not be any unobserved factors such as ambition or work ethic that influence both whether a person enrolls in the program and the wage received upon completion. To use the methods implemented by

the `teffects` estimators, we must have variables in our dataset that allow us to control for those types of factors.

We mentioned the third assumption, overlap, in our discussions of IPW. More formally, the overlap assumption states that each individual have a positive probability of receiving treatment. In our birthweight example, we noted that there were no observations on young smokers and older nonsmokers. Perhaps we just have an unlucky sample, but to accurately assess the impact of treatment using these methods, we must have overlap to accurately estimate the counterfactual birthweights. In the context of matching estimators, overlap essentially means that we can actually match treated subjects with similar nontreated subjects.

A quick tour of the estimators

The `teffects` command implements six estimators of treatment effects. We introduce each one by showing the basic syntax one would use to apply them to our birthweight example. See each command's entry for more information.

Regression adjustment

`teffects ra` implements the RA estimator. We estimate the effect of a mother's smoking behavior (`mbsmoke`) on the birthweight of her child (`bweight`), controlling for marital status (`mmarried`), the mother's age (`mage`), whether the mother had a prenatal doctor's visit in the baby's first trimester (`prenatal1`), and whether this baby is the mother's first child (`fbaby`). We use linear regression (the default) to model `bweight`:

```
. use http://www.stata-press.com/data/r13/cattaneo2
. teffects ra (bweight mmarried mage prenatal1 fbaby) (mbsmoke)
```

Inverse-probability weighting

`teffects ipw` implements the IPW estimator. Here we estimate the effect of smoking by using a probit model to predict the mother's smoking behavior as a function of marital status, the mother's age, and indicators for first-trimester doctor's visits and firstborn status:

```
. teffects ipw (bweight) (mbsmoke mmarried mage prenatal1 fbaby, probit)
```

Inverse-probability-weighted regression adjustment

`teffects ipwra` implements the IPWRA estimator. We model the outcome, birthweight, as a linear function of marital status, the mother's age, and indicators for first-trimester doctor's visits and firstborn status. We use a logistic model (the default) to predict the mother's smoking behavior, using the same covariates as explanatory variables:

```
. teffects ipwra (bweight mmarried mage prenatal1 fbaby) ///
(mbsmoke mmarried mage prenatal1 fbaby)
```

Augmented inverse-probability weighting

`teffects aipw` implements the AIPW estimator. Here we use the same outcome- and treatment-model specifications as we did with the IPWRA estimator:

```
. teffects aipw (bweight mmarried mage prenatal1 fbaby) ///  
              (mbsmoke mmarried mage prenatal1 fbaby)
```

Nearest-neighbor matching

`teffects nnmatch` implements the NNM estimator. In this example, we match treated and untreated subjects based on marital status, the mother's age, the father's age, and indicators for first-trimester doctor's visits and firstborn status. We use the Mahalanobis distance based on the mother's and father's ages to find matches. We use exact matching on the other three variables to enforce the requirement that treated subjects are matched with untreated subjects who have the same marital status and indicators for first-trimester doctor's visits and firstborn statuses. Because we are matching on two continuous covariates, we request that `teffects nnmatch` include a bias-correction term based on those two covariates:

```
. teffects nnmatch (bweight mage fage) (mbsmoke),          ///  
                  ematch(prenatal1 mmarried fbaby) biasadj(mage fage)
```

Propensity-score matching

`teffects psmatch` implements the PSM estimator. Here we model the propensity score using a probit model, incorporating marital status, the mother's age, and indicators for first-trimester doctor's visits and firstborn status as covariates:

```
. teffects psmatch (bweight) (mbsmoke mmarried mage prenatal1 fbaby, probit)
```

Video example

[Introduction to treatment effects in Stata, part 1](#)

Reference

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

Also see

[TE] [teffects](#) — Treatment-effects estimation for observational data

[TE] [teffects intro advanced](#) — Advanced introduction to treatment effects for observational data