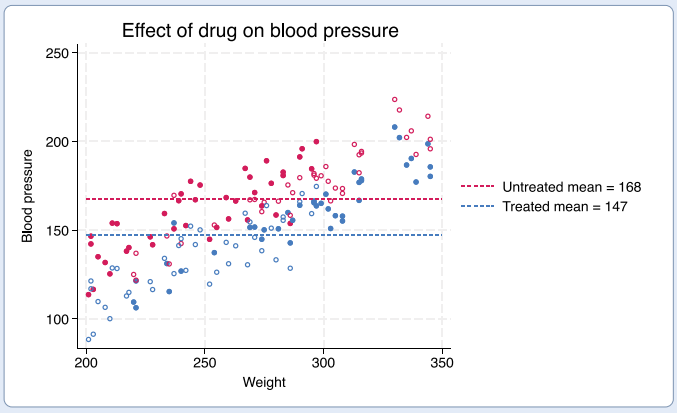
Causal inference

Propensity-score matching, IPW, and more

Stata's features for causal inference allow you to estimate experimental-type causal effects (treatment effects) from observational data. Are you interested in a continuous, binary, count, or survival outcome? Are you modeling the outcome process or treatment process? Stata can estimate your treatment effect. With such a comprehensive suite of estimators, you will find the one that's right for you.



Estimators

- Difference in differences (DID)
- Heterogeneous DID
- Inverse-probability weights (IPW)
- Propensity-score matching
- Covariate matching
- Regression adjustment
- Weighted regression
- Doubly robust methods
  - Augmented IPW (AIPW)
  - IPW with regression adjustment
  - AIPW with lasso selection of controls
- Causal mediation
- Conditional average treatment effects (CATEs) New

Statistics

- Average treatment effects (ATEs)
- ATEs on the treated (ATETs)
- Potential-outcome means (POMs)
- Direct effects, indirect effects
- Individualized ATEs (IATEs) New
- Group ATEs (GATEs) New
- Sorted group ATEs (GATESs) New

Outcomes

- Continuous—linear
- Binary—logistic, probit, heteroskedastic probit
- Count—Poisson
- Fractional
- Nonnegative, including exponential mean
- Survival—exponential, Weibull, gamma, lognormal

Treatments

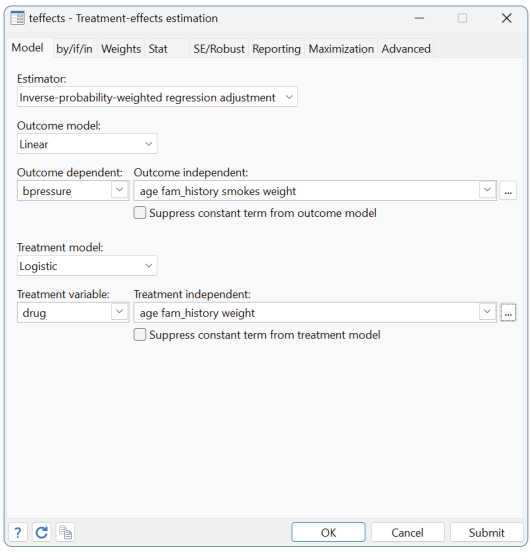
- Binary—logistic, probit, heteroskedastic probit
- Multivalued—multinomial logistic

Diagnostics

- Overlap plots
- Covariate balance

Endogenous treatment effects

- Continuous, censored, binary, ordinal, and count outcomes
- ATEs, ATETs, and POMs
- Combine with endogenous covariates, sample selection, and panel data
- Test for endogeneity



Perhaps you are a medical researcher who knows the variables that doctors consider when deciding whether to prescribe a drug for treatment of high blood pressure, but you are far less confident about the variables that should be used to model blood pressure.

You might use the IPW estimator,

```
. teffects ipw (bp) (drug x1 x2)
```

or propensity-score matching,

```
. teffects psmatch (bp) (drug x1 x2)
```

Know more about variables that impact blood pressure but not about those that determine whether a drug is prescribed?

Use regression adjustment,

```
. teffects ra (bp x1 x3) (drug)
```

or nearest-neighbor matching,

```
. teffects nnmatch (bp x1 x3) (drug)
```

If you know something about modeling both blood pressure and whether the drug is prescribed, you can use one of the doubly robust estimators.

Use augmented IPW,

```
. teffects aipw (bp x1 x3) (drug x1 x2)
```

or IPW with regression adjustment,

```
. teffects ipwra (bp x1 x3) (drug x1 x2)
```

Surprisingly, with these doubly robust methods, we need to be right about only one of the two model specifications.

Let's see an example using IPW with regression adjustment. We model systolic blood pressure as a function of age, family history of high blood pressure, smoking, and weight. Whether a drug is prescribed or not is modeled as a function of age, family history, and weight.

```
. teffects ipwra
            (bp age fam_history smokes weight)
            (drug age fam_history weight)
```

Viewer - view te1.smcl

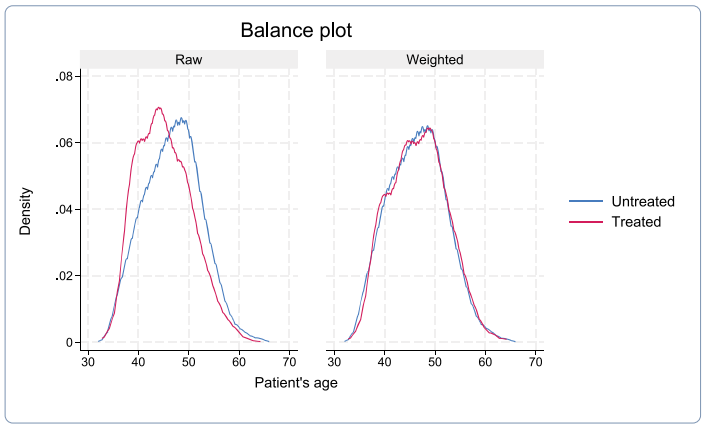
```
Treatment-effects estimation                   Number of obs     =      4,642
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: logit
```

bp	Coefficient	Robust std. err.	z	P>|z|	[95% conf. interval]	
ATE						
drug (Treated vs Untreated)	-21.38595	.6369856	-33.57	0.000	-22.63442	-20.13748
POmean						
drug Untreated	167.5199	.2507166	668.16	0.000	167.0285	168.0113

The ATE is –21.4. If all patients were prescribed the drug, average blood pressure would decrease by 21.4 mmHg. The POM of 167.5 gives us the estimated systolic blood pressure if no one were given the drug.

Let's check some diagnostics. Are the covariates balanced? We can look at the kernel density plots of age, for instance, comparing the treated and untreated.

```
. tebalance density age
```



After weighting, the densities for age are very similar.

We can test for balancing of all covariates.

```
. tebalance overid
```

Viewer - view te2.smcl

```
Overidentification test for covariate balance
H0: Covariates are balanced

        chi2(7)      =  10.1924
        Prob > chi2  =   0.1779
```

We do not reject the null hypothesis that all covariates are balanced.

We have only touched the extent of treatment-effects estimators that are available.

Do you have a binary outcome?

```
. teffects ipwra (y x1 x2, probit) (treat x1 x3)
```

Or a survival outcome?

```
. stteffects ipwra (x1 x2) (treat x1 x3)
```

Do you have an endogeneity problem?

```
. eteffects (y x1 x2) (treat x1 x3)
```

Do you have many potential controls?

```
. telasso (y x1 x2 x3 x4 ...)
          (treat x1 x2 x3 x4 ...)
```

Do you have a mediator?

```
. mediate (y x1 x2) (m x1 x3) (treat)
```

Do different groups have different ATEs?

```
. cate po (y x1 i.x2) (treat), group(x2)
```


© 2025 StataCorp LLC
Stata is a registered trademark of StataCorp LLC
4905 Lakeway Drive, College Station, TX 77845, USA.