

Introduction to causal inference and treatment effects

July 9th, 2025

Topics

Part I: Concepts and motivation

- Potential outcomes
- The fundamental problem of causal inference
- The average treatment effect
- Experimental vs observational data

Part II: The **teffects** suite of commands

- Outcome models with **teffects ra**
- Treatment models with **teffects ipw**
- Doubly-robust estimation with **teffects ipwra** and **teffects aipw**
- Matching with **teffects nnmatch** and **teffects psmatch**
- Postestimation tools
- Bonus: heterogeneous effects with **cate**
- Bonus: covariate selection with **telasso**
- Bonus: survival analysis with **stteffects**
- Bonus: mediation with **mediate**

Part I: Concepts and motivation

Outcome ↓ Y	Binary treatment ↓ T	Outcome under control ↓ Y(0)	Outcome under treatment ↓ Y(1)	Covariates	
				X ₁	X ₂
3.1	1	?	3.1	#	#
4.2	1	?	4.2	#	#
2.9	1	?	2.9	#	#
1.7	0	1.7	?	#	#
2.1	0	2.1	?	#	#
3.8	0	3.8	?	#	#
...					
...					

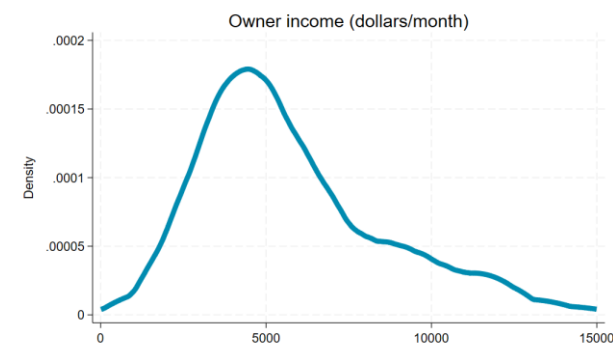
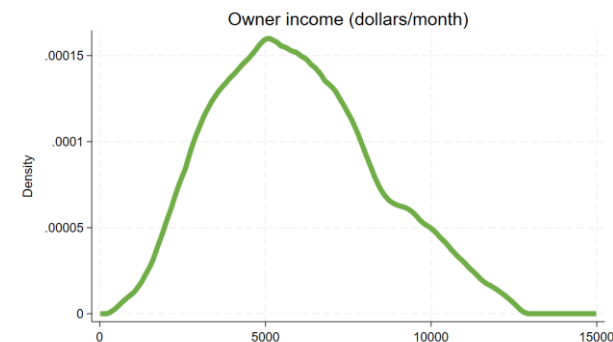
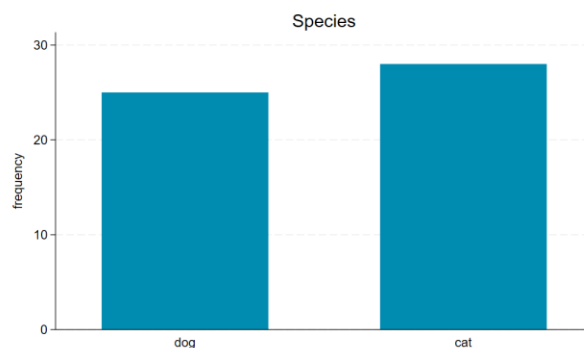
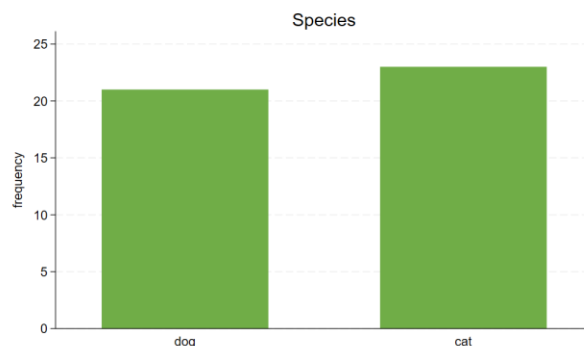
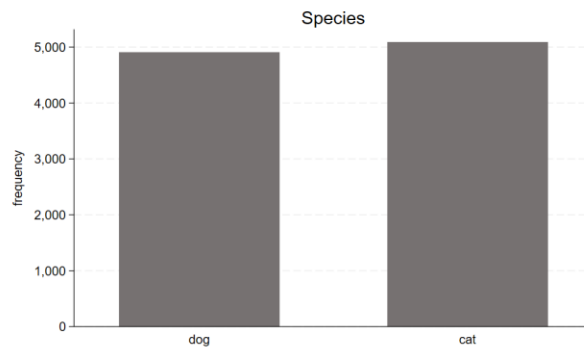
Y	Y(0)	Y(1)
3.1	?	3.1
4.2	?	4.2
2.9	?	2.9
1.7	1.7	?
2.1	2.1	?
3.8	3.8	?

•
•
•

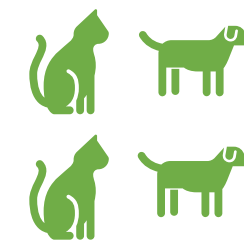
$$ATE = E[Y(1)] - E[Y(0)]$$

Y	Y(0)	Y(1)
3.1	?	3.1
4.2	?	4.2
2.9	?	2.9
1.7	1.7	?
2.1	2.1	?
3.8	3.8	?

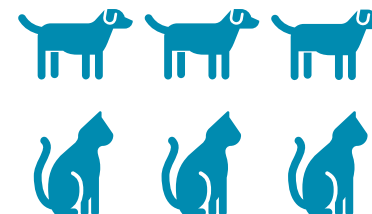
⋮



Population



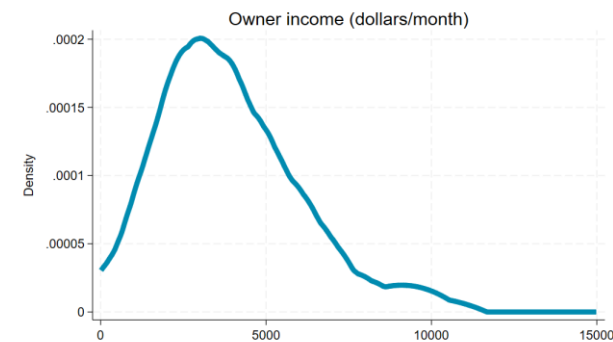
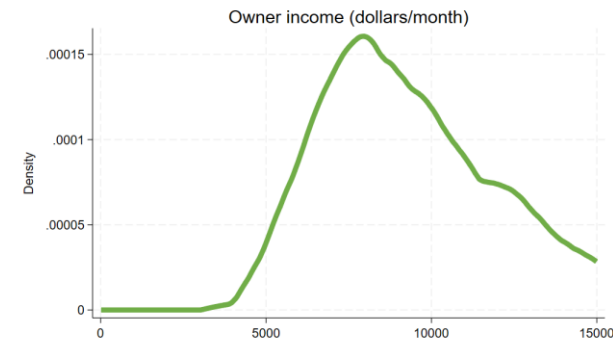
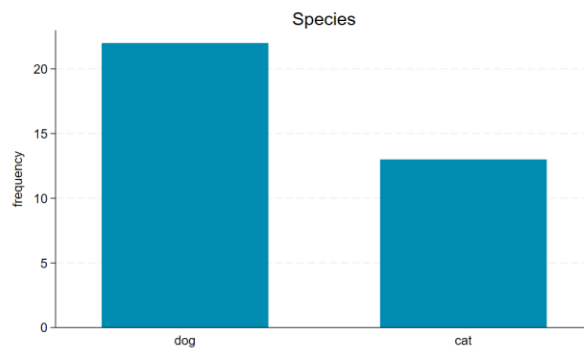
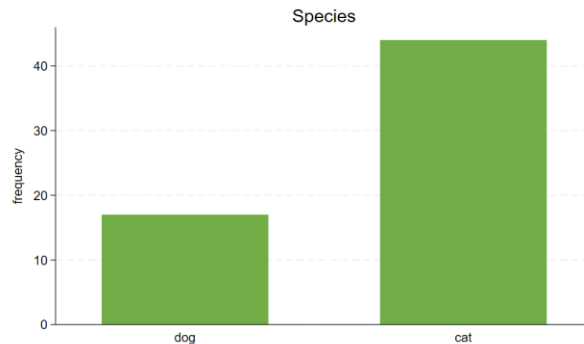
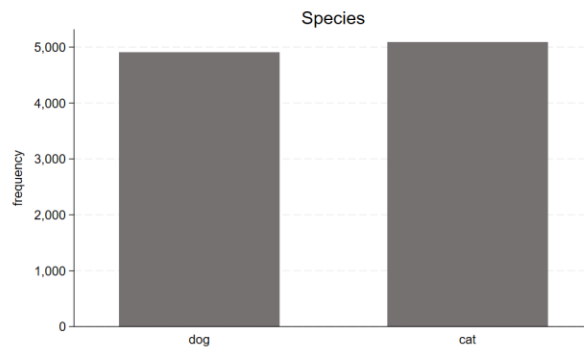
Treated



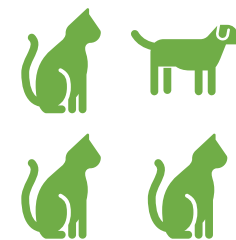
Controls

Y	Y(0)	Y(1)
3.1	?	3.1
4.2	?	4.2
2.9	?	2.9
1.7	1.7	?
2.1	2.1	?
3.8	3.8	?

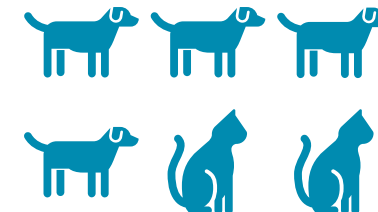
⋮



Population

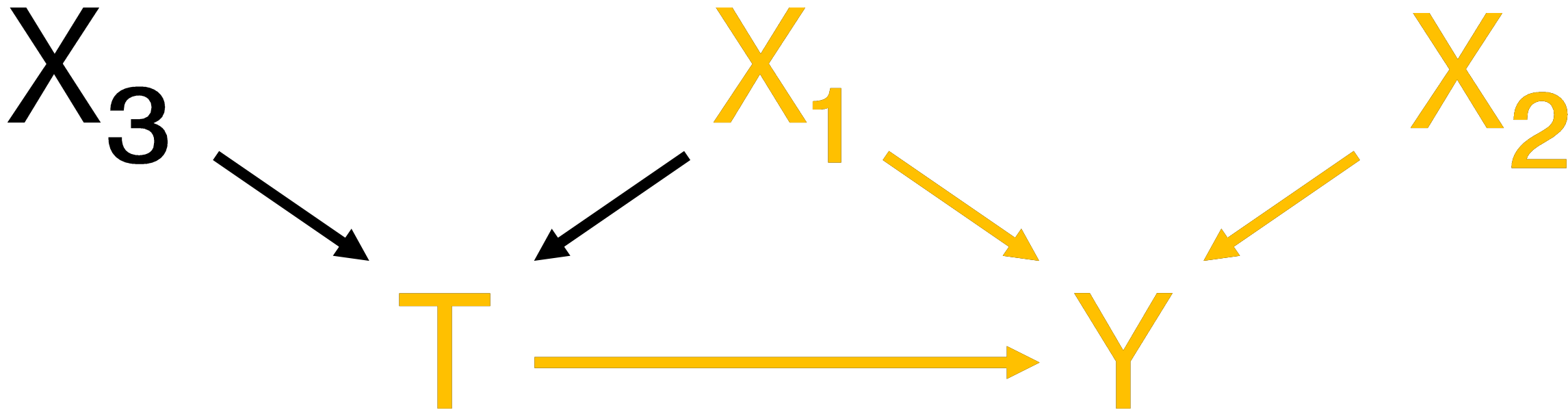


Treated



Controls

Part II: The teffects suite of commands



$$Y = f(T, X_1, X_2) + \epsilon_Y$$

Viewer - help teffects

File Edit History Help

help teffects

help teffects x

Dialog Also see Jump to

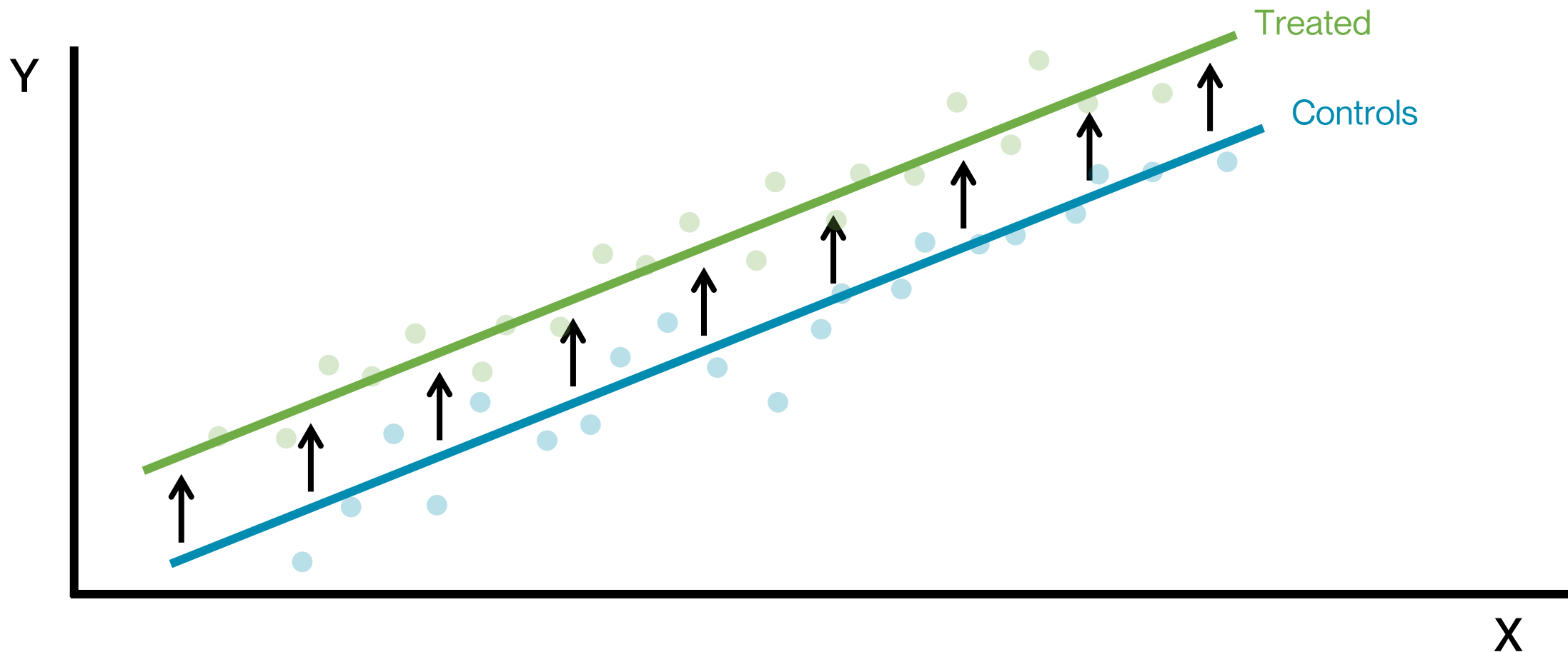
[CAUSAL] teffects — Treatment-effects estimation for observational data
([View complete PDF manual entry](#))

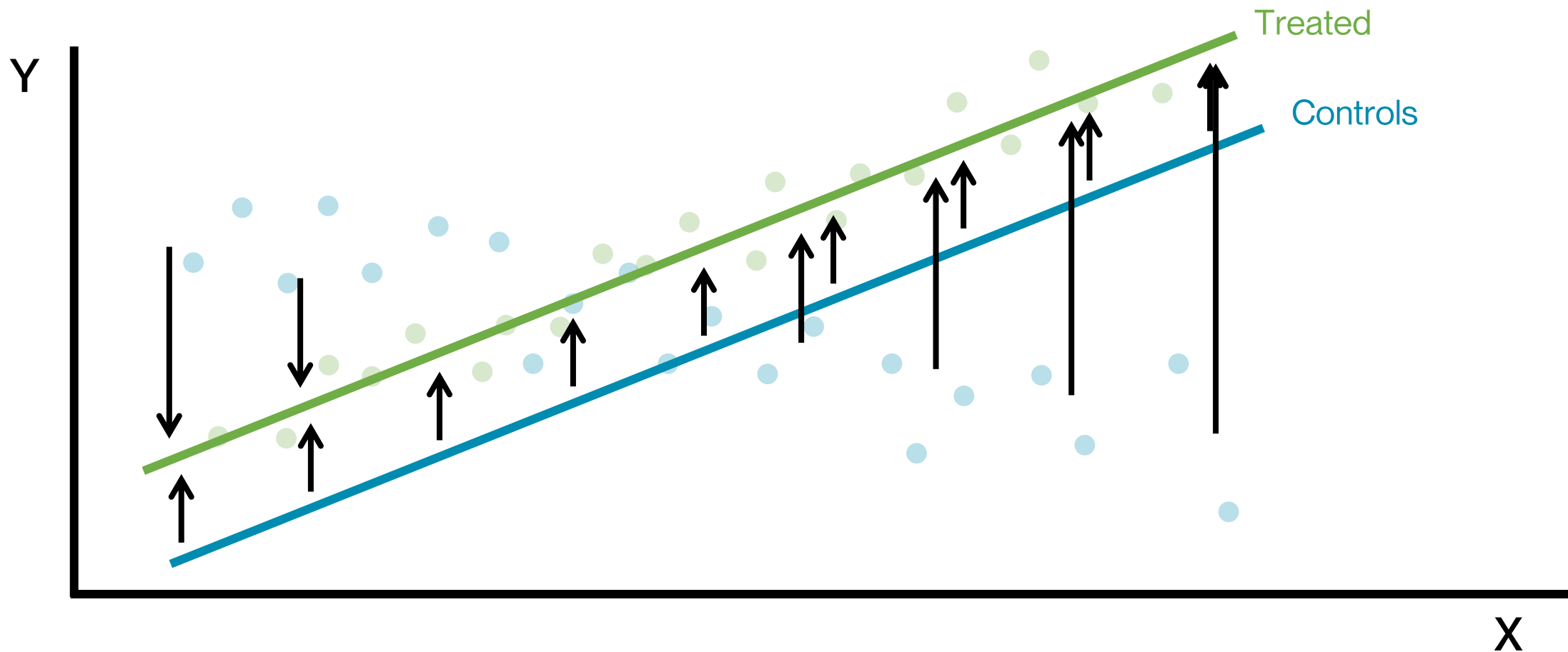
Syntax

teffects *subcommand* ... [, *options*]

<i>subcommand</i>	Description
aipw	augmented inverse-probability weighting
ipw	inverse-probability weighting
ipwra	inverse-probability-weighted regression adjustment
nnmatch	nearest-neighbor matching
psmatch	propensity-score matching
ra	regression adjustment

CAP NUM INS





Variables Y ⌵ ×

 Filter variables here

Name	Label
id	Identification code
low	Birthweight<2500g
age	Age of mother
lwt	Weight at last menstrual period
race	Race
smoke	Smoked during pregnancy
ptl	Premature labor history (count)
ht	Has history of hypertension
ui	Presence, uterine irritability
ftv	Number of visits to physician during 1st trim...
bwt	Birthweight (grams)

```
. webuse lbw
```

Treatment-effects estimation Number of obs = 189
 Estimator : **regression adjustment**
 Outcome model : **linear**
 Treatment model : **none**

	bwt	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	smoke						
(Smoker vs Nonsmoker)		-242.198	103.4367	-2.34	0.019	-444.9303	-39.4657
POmean							
	smoke						
Nonsmoker		3020.813	73.21419	41.26	0.000	2877.316	3164.31

```
. teffects ra (bwt age ui ptl) (smoke)
```

Treatment-effects estimation Number of obs = 189
 Estimator : regression adjustment
 Outcome model : linear
 Treatment model : none

	bwt	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	smoke (Smoker vs Nonsmoker)	-242.198	103.4367	-2.34	0.019	-444.9303	-39.4657
POMean							
	smoke Nonsmoker	3020.813	73.21419	41.26	0.000	2877.316	3164.31
OME0							
	age	29.81673	12.65627	2.36	0.018	5.010903	54.62255
	ui	-659.1187	218.8562	-3.01	0.003	-1088.069	-230.1684
	ptl	-227.7906	242.3873	-0.94	0.347	-702.8609	247.2798
	_cons	2470.17	293.4731	8.42	0.000	1894.974	3045.367
OME1							
	age	-27.22703	17.43666	-1.56	0.118	-61.40225	6.948196
	ui	-484.9099	210.1909	-2.31	0.021	-896.8765	-72.9433
	ptl	-8.030069	155.4444	-0.05	0.959	-312.6954	296.6353
	_cons	3484.73	411.8119	8.46	0.000	2677.593	4291.866

```
. teffects ra (bwt age ui ptl) (smoke), aequations
```

Treatment-effects estimation Number of obs = 189
 Estimator : **regression adjustment**
 Outcome model : **linear**
 Treatment model : **none**

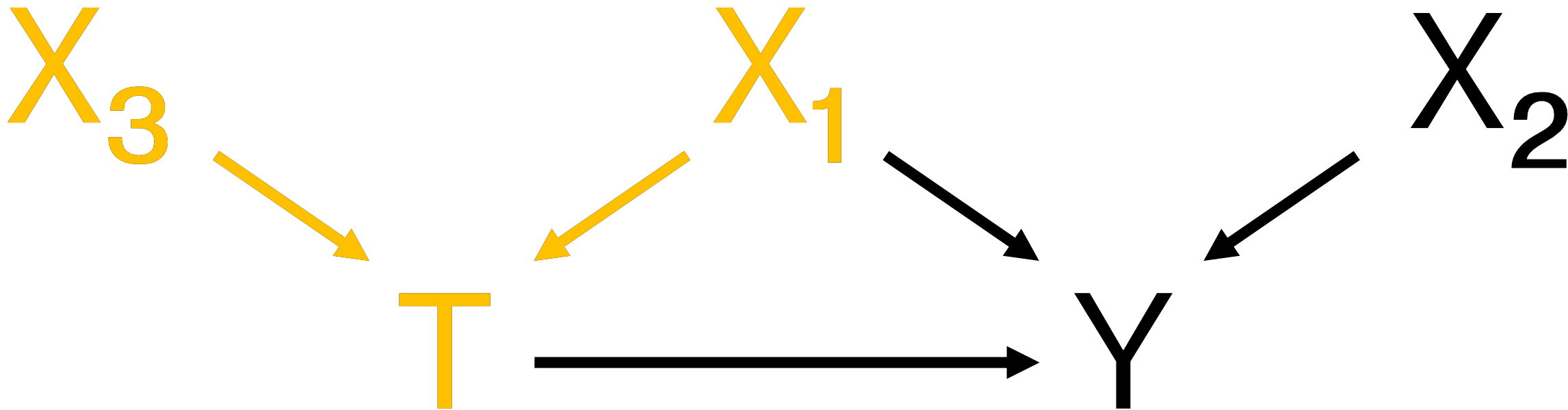
bwt		Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATET	smoke						
(Smoker vs Nonsmoker)		-195.4551	119.916	-1.63	0.103	-430.4862	39.57594
POmean	smoke						
Nonsmoker		2967.752	97.55975	30.42	0.000	2776.539	3158.966

```
. teffects ra (bwt age ui ptl) (smoke), atet
```

Treatment-effects estimation Number of obs = 189
 Estimator : **regression adjustment**
 Outcome model : **linear**
 Treatment model : **none**

bwt	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
PMeans						
smoke						
Nonsmoker	3020.813	73.21419	41.26	0.000	2877.316	3164.31
Smoker	2778.615	74.72794	37.18	0.000	2632.151	2925.079

```
. teffects ra (bwt age ui ptl) (smoke), pom
```



$$T = g(X_1, X_3) + \epsilon_T$$

Viewer - help teffects

File Edit History Help

help teffects

help teffects x

Dialog Also see Jump to

[CAUSAL] teffects — Treatment-effects estimation for observational data
([View complete PDF manual entry](#))

Syntax

teffects *subcommand* ... [, *options*]

<i>subcommand</i>	Description
aipw	augmented inverse-probability weighting
ipw	inverse-probability weighting
ipwra	inverse-probability-weighted regression adjustment
nnmatch	nearest-neighbor matching
psmatch	propensity-score matching
ra	regression adjustment

CAP NUM INS

Y	Y(0)	Y(1)	T	p	w
3.1	?	3.1	1	#	1/p
4.2	?	4.2	1	#	1/p
2.9	?	2.9	1	#	1/p
1.7	1.7	?	0	#	1/(1-p)
2.1	2.1	?	0	#	1/(1-p)
3.8	3.8	?	0	#	1/(1-p)

• • • •
• • • •
• • • •

$$\widehat{ATE} = \hat{E}[Y(1)] - \hat{E}[Y(0)]$$

Treatment-effects estimation Number of obs = 189
 Estimator : **inverse-probability weights**
 Outcome model : **weighted mean**
 Treatment model: **logit**

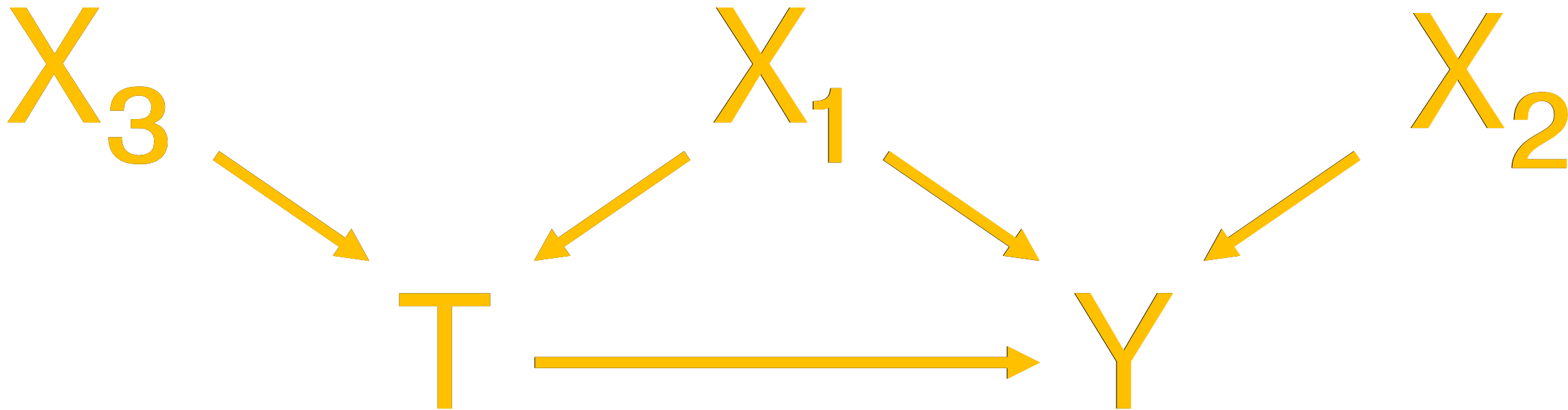
	bwt	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	smoke						
(Smoker vs Nonsmoker)		-279.9708	102.4713	-2.73	0.006	-480.8108	-79.13082
P0mean							
	smoke						
Nonsmoker		3050.078	68.74209	44.37	0.000	2915.346	3184.81

```
. teffects ipw (bwt) (smoke age ftv)
```

Treatment-effects estimation Number of obs = 189
 Estimator : **inverse-probability weights**
 Outcome model : **weighted mean**
 Treatment model: **logit**

	bwt	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	smoke (Smoker vs Nonsmoker)	-279.9708	102.4713	-2.73	0.006	-480.8108	-79.13082
POmean							
	smoke Nonsmoker	3050.078	68.74209	44.37	0.000	2915.346	3184.81
TME1							
	age	-.0157273	.0284615	-0.55	0.581	-.0715109	.0400563
	ftv	-.0380275	.155715	-0.24	0.807	-.3432233	.2671683
	_cons	-.0463427	.6640162	-0.07	0.944	-1.347791	1.255105

```
. teffects ipw (bwt) (smoke age ftv), aequations
```



$$T = g(X_1, X_3) + \epsilon_T \quad Y = f(T, X_1, X_2) + \epsilon_Y$$

Viewer - help teffects

File Edit History Help

help teffects

help teffects x

Dialog Also see Jump to

[CAUSAL] teffects — Treatment-effects estimation for observational data
([View complete PDF manual entry](#))

Syntax

teffects *subcommand* ... [, *options*]

<i>subcommand</i>	Description
aipw	augmented inverse-probability weighting
ipw	inverse-probability weighting
ipwra	inverse-probability-weighted regression adjustment
nnmatch	nearest-neighbor matching
psmatch	propensity-score matching
ra	regression adjustment

CAP NUM INS

Treatment-effects estimation
 Estimator : **augmented IPW**
 Outcome model : **linear by ML**
 Treatment model: **logit**

Number of obs = 189

bwt		Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE	smoke						
	(Smoker vs Nonsmoker)	-239.6607	103.1649	-2.32	0.020	-441.8601	-37.46131
P0mean	smoke						
	Nonsmoker	3020.753	73.35362	41.18	0.000	2876.983	3164.524

```
. teffects aipw (bwt age ui ptl) (smoke age ftv)
```

Treatment-effects estimation
 Estimator : **augmented IPW**
 Outcome model : **linear by ML**
 Treatment model: **logit**

Number of obs = 189

	bwt	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	smoke (Smoker vs Nonsmoker)	-239.6607	103.1649	-2.32	0.020	-441.8601	-37.46131
P0mean							
	smoke Nonsmoker	3020.753	73.35362	41.18	0.000	2876.983	3164.524
OME0							
	age	29.81673	12.65627	2.36	0.018	5.010903	54.62255
	ui	-659.1187	218.8562	-3.01	0.003	-1088.069	-230.1684
	ptl	-227.7906	242.3873	-0.94	0.347	-702.8609	247.2798
	_cons	2470.17	293.4731	8.42	0.000	1894.974	3045.367
OME1							
	age	-27.22703	17.43666	-1.56	0.118	-61.40225	6.948196
	ui	-484.9099	210.1909	-2.31	0.021	-896.8765	-72.9433
	ptl	-8.030069	155.4444	-0.05	0.959	-312.6954	296.6353
	_cons	3484.73	411.8119	8.46	0.000	2677.593	4291.866
TME1							
	age	-.0157273	.0284615	-0.55	0.581	-.0715109	.0400563
	ftv	-.0380275	.155715	-0.24	0.807	-.3432233	.2671683
	_cons	-.0463427	.6640162	-0.07	0.944	-1.347791	1.255105

```
. teffects aipw (bwt age ui ptl) (smoke age ftv), aequations
```

Viewer - help teffects

File Edit History Help

help teffects

help teffects x

Dialog Also see Jump to

[CAUSAL] teffects — Treatment-effects estimation for observational data
([View complete PDF manual entry](#))

Syntax

teffects *subcommand* ... [, *options*]

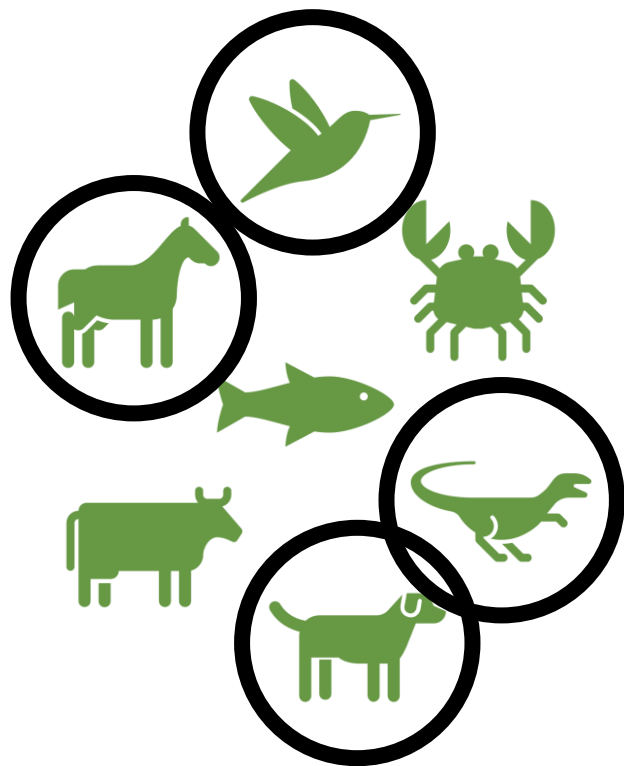
<i>subcommand</i>	Description
aipw	augmented inverse-probability weighting
ipw	inverse-probability weighting
ipwra	inverse-probability-weighted regression adjustment
nnmatch	nearest-neighbor matching
psmatch	propensity-score matching
ra	regression adjustment

CAP NUM INS

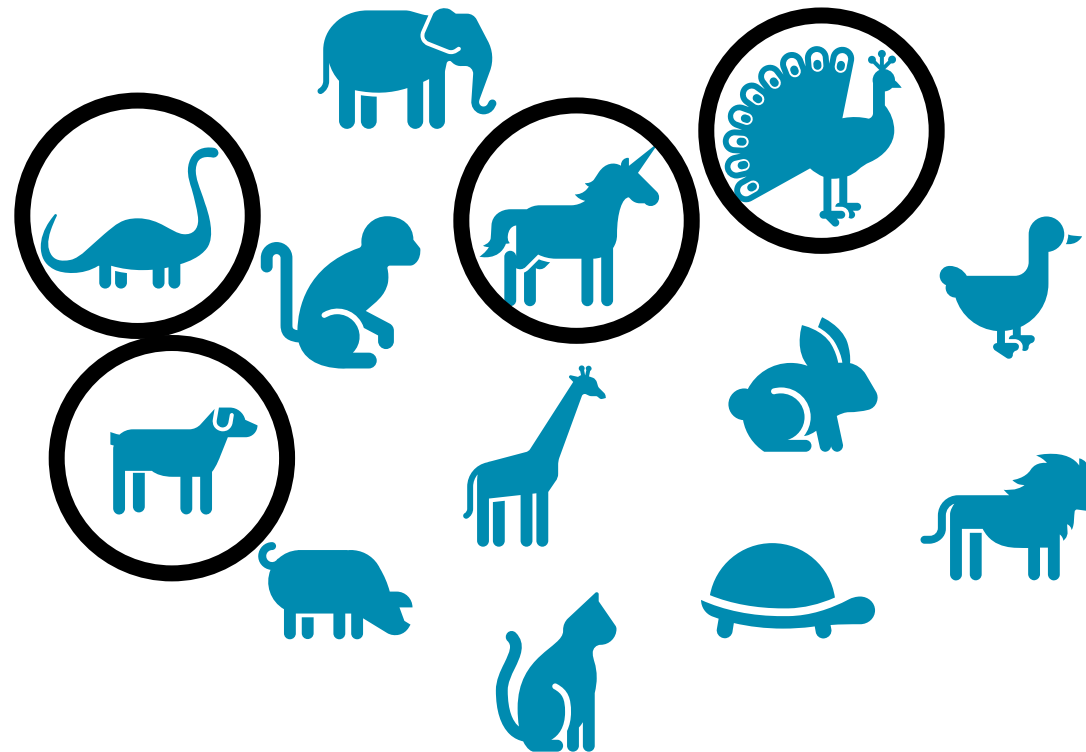
Y	Y(0)	Y(1)
3.1	?	3.1
4.2	?	4.2
2.9	?	2.9
1.7	1.7	?
2.1	2.1	?
3.8	3.8	?

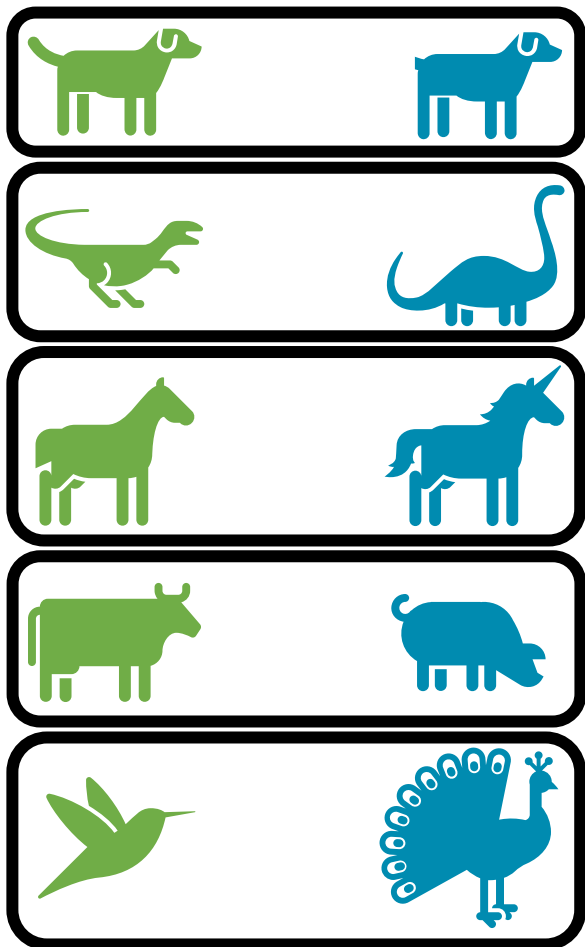
•
•
•

Treated



Controls





teffects nnmatch

- Which distance metric?
- Multiple matches?
- Caliper?
- Exact matching?

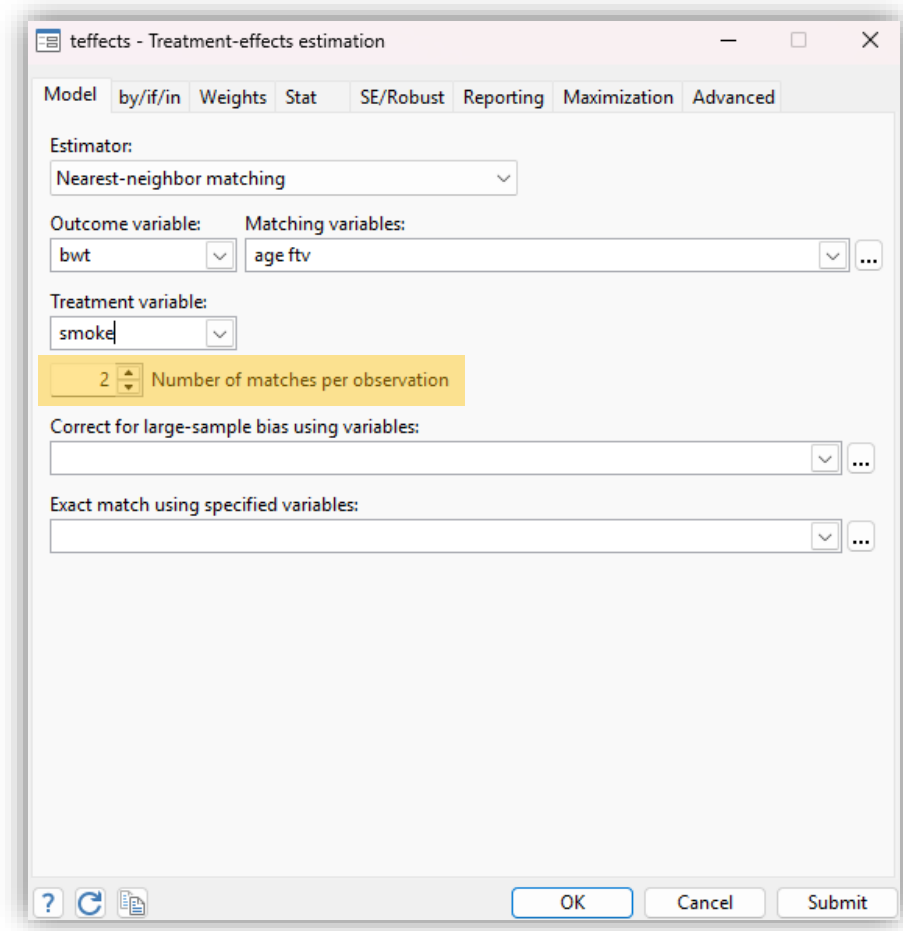
teffects psmatch

- Multiple matches?
- Caliper?

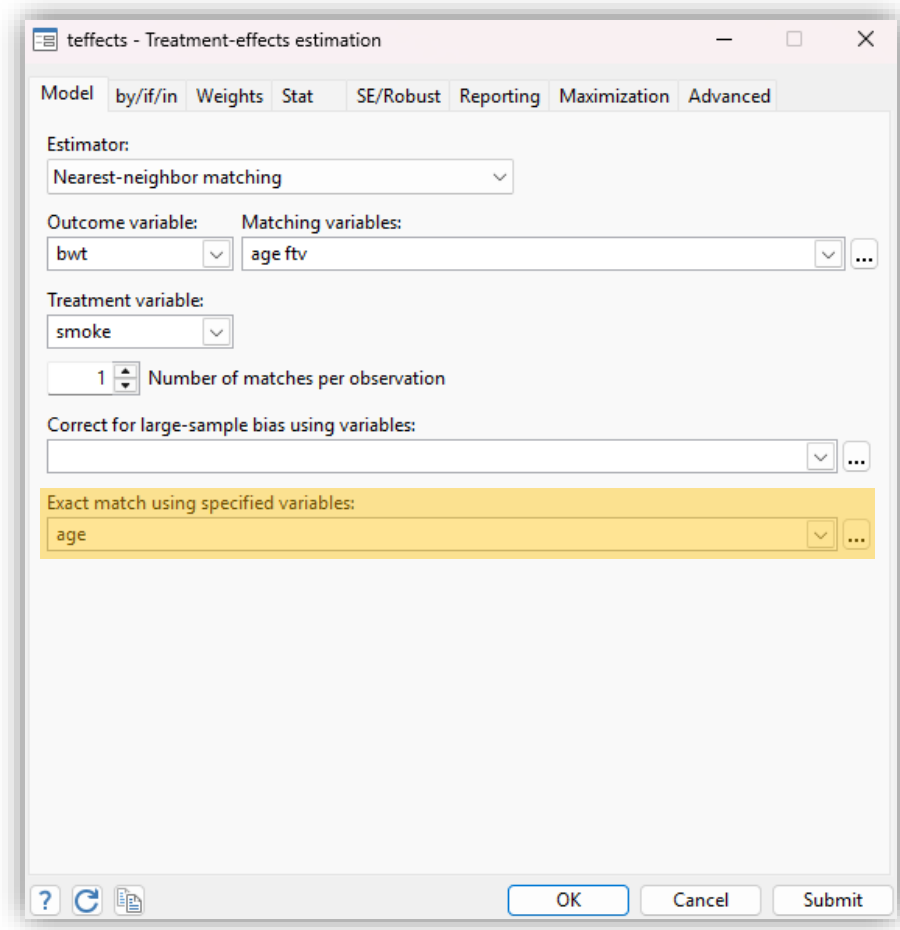
Treatment-effects estimation	Number of obs	=	189
Estimator : nearest-neighbor matching	Matches: requested	=	1
Outcome model : matching	min	=	1
Distance metric: Mahalanobis	max	=	7

bwt		Coefficient	AI robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	smoke						
(Smoker vs Nonsmoker)		-315.9859	108.5831	-2.91	0.004	-528.8049	-103.167

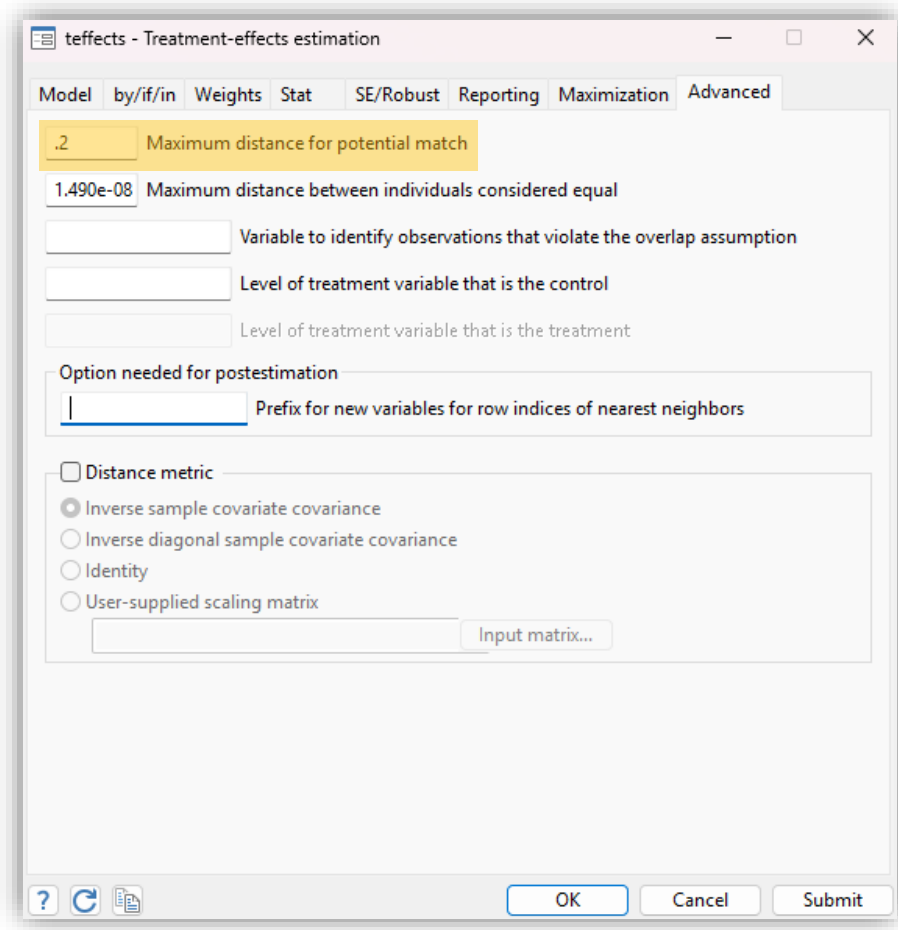
```
. teffects nnmatch (bwt age ftv) (smoke)
```



```
. teffects nnmatch (bwt age ftv) (smoke), nneighbor(2)
```



```
. teffects nnmatch (bwt age ftv) (smoke), ematch(age)
```



```
. teffects nnmatch (bwt age ftv) (smoke), caliper(.2)
```

teffects - Treatment-effects estimation

Model by/if/in Weights Stat SE/Robust Reporting Maximization Advanced

Maximum distance for potential match

1.490e-08 Maximum distance between individuals considered equal

Variable to identify observations that violate the overlap assumption

Level of treatment variable that is the control

Level of treatment variable that is the treatment

Option needed for postestimation

Prefix for new variables for row indices of nearest neighbors

☒ Distance metric

☐ Inverse sample covariate covariance

☐ Inverse diagonal sample covariate covariance

☒ Identity

☐ User-supplied scaling matrix

Input matrix...

? ↺ 📄 OK Cancel Submit

```
. teffects nnmatch (bwt age ftv) (smoke), metric(euclidean)
```

Treatment-effects estimation	Number of obs	=	189
Estimator : propensity-score matching	Matches: requested	=	1
Outcome model : matching	min	=	1
Treatment model: logit	max	=	7

bwt		Coefficient	AI robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	smoke						
(Smoker vs Nonsmoker)		-318.7534	115.9251	-2.75	0.006	-545.9624	-91.54439

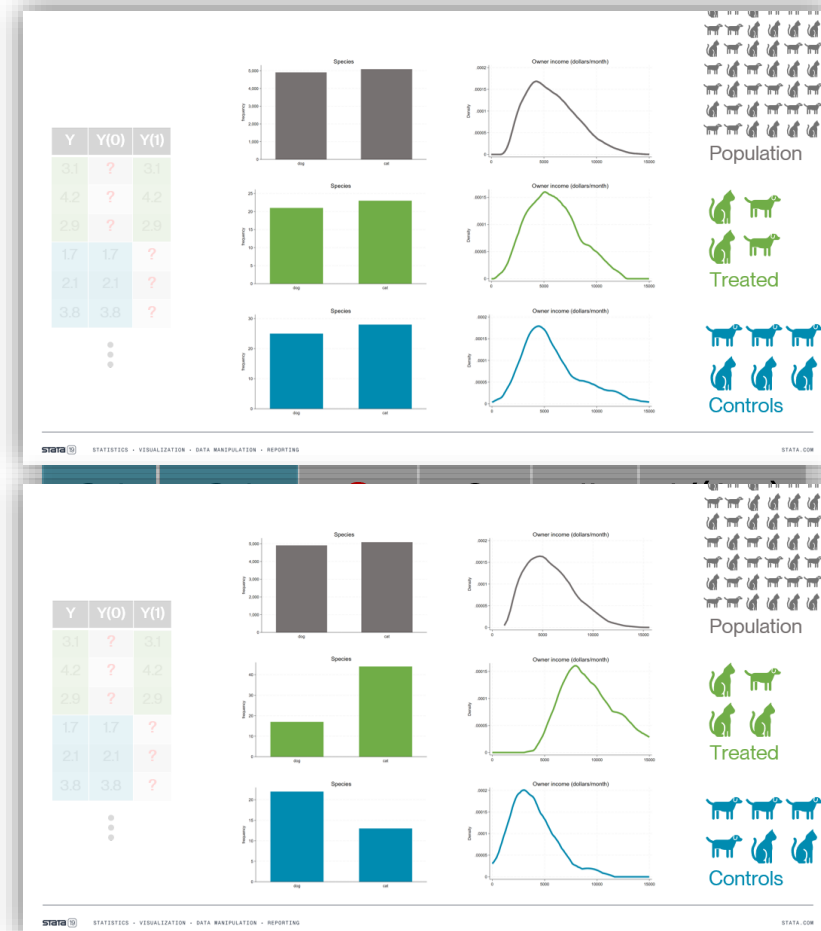
```
. teffects psmatch (bwt) (smoke age ftv)
```

Command	Estimand	Outcome types	Treatment types	Models specified
<code>teffects ra</code>	ATE ATET POM	continuous binary count fractional nonnegative	binary multivalued	outcome
<code>teffects ipw</code>	ATE ATET POM	continuous binary count fractional nonnegative	binary multivalued	treatment
<code>teffects ipwra</code>	ATE ATET POM	continuous binary count fractional nonnegative	binary multivalued	outcome treatment
<code>teffects aipw</code>	ATE ATET POM	continuous binary count fractional nonnegative	binary multivalued	outcome treatment
<code>teffects psmatch</code>	ATE ATET	continuous binary count fractional nonnegative	binary	treatment
<code>teffects nnmatch</code>	ATE ATET	continuous binary count fractional nonnegative	binary	outcome*
* <code>nnmatch</code> includes covariates for modeling the outcome but does not require specification of a functional form for the outcome model.				

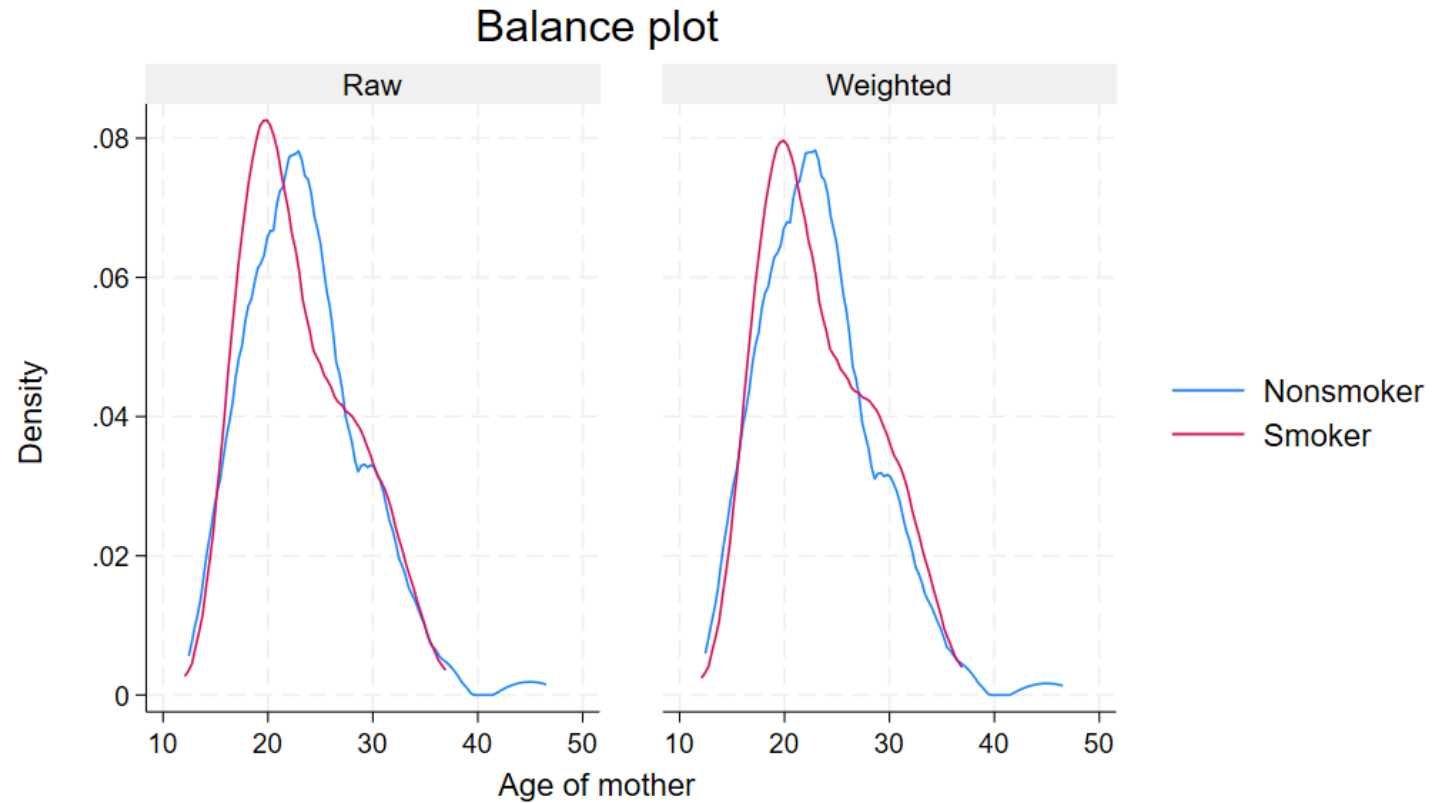
Covariate balance summary

	Raw	Weighted
Number of obs =	189	189.0
Treated obs =	74	94.5
Control obs =	115	94.5

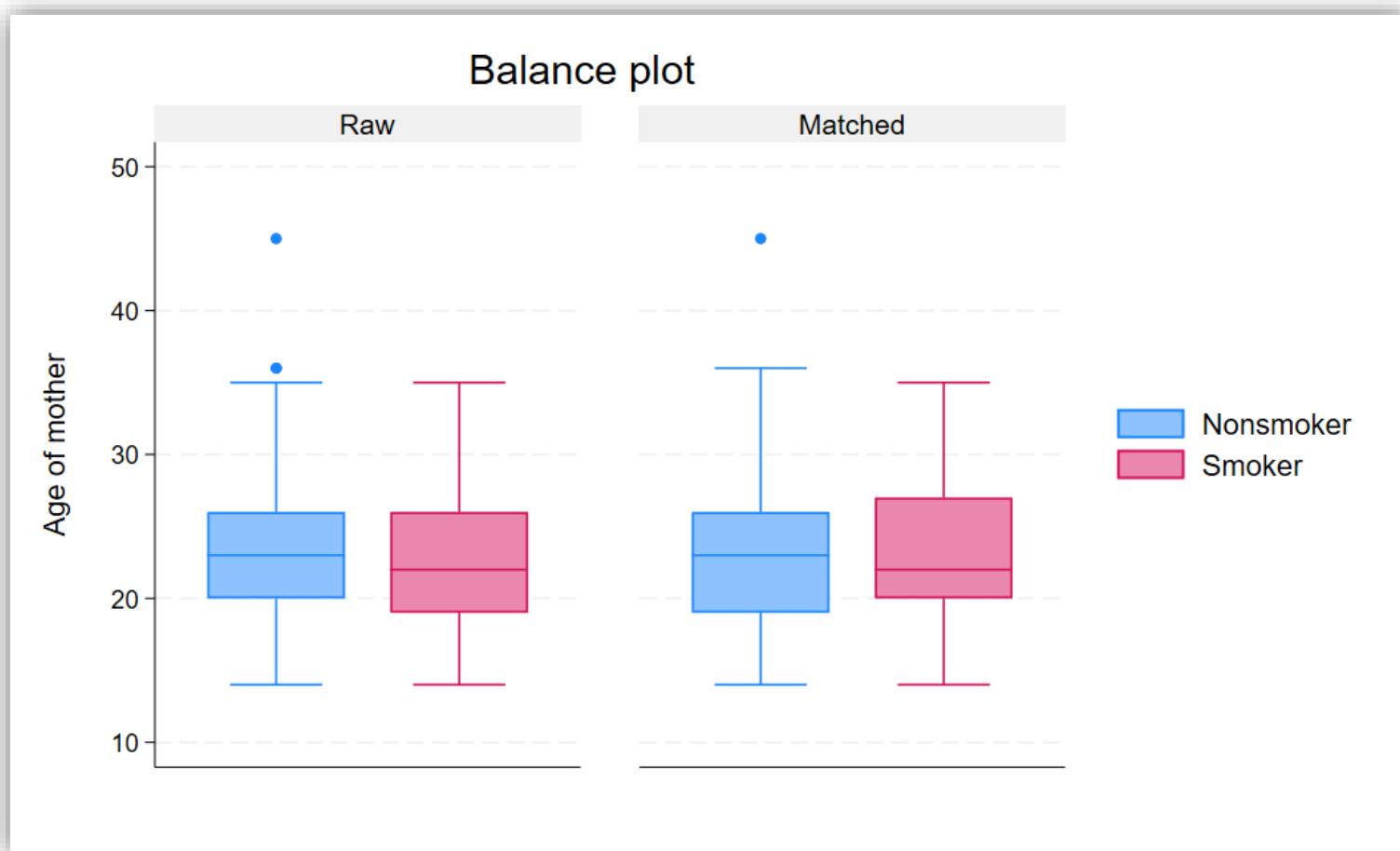
	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
age	-.091251	-.0008564	.8521761	.8967359
ftv	-.0559412	.0080289	1.45373	1.585167



- . teffects ipw (bwt) (smoke age ftv)
- . tebalance summarize



```
. teffects ipw (bwt) (smoke age ftv)  
. tebalance density age
```



```
. teffects psmatch (bwt) (smoke age ftv)  
. tebalance box age
```

Overidentification test for covariate balance
H0: Covariates are balanced

```
chi2(3)      = .730127  
Prob > chi2   = 0.8661
```

```
. teffects aipw (bwt age ui ptl) (smoke age ftv)  
. tebalance overid
```

Viewer - help cate

File Edit History Help

help cate

help cate

[CAUSAL] cate — Conditional average treatment-effects estimation
([View complete PDF manual entry](#))

Syntax

Partialing-out estimator

```
cate po (ovar catevarlist) (tvar) [if] [in] [, options]
```

Augmented inverse-probability weighting estimator

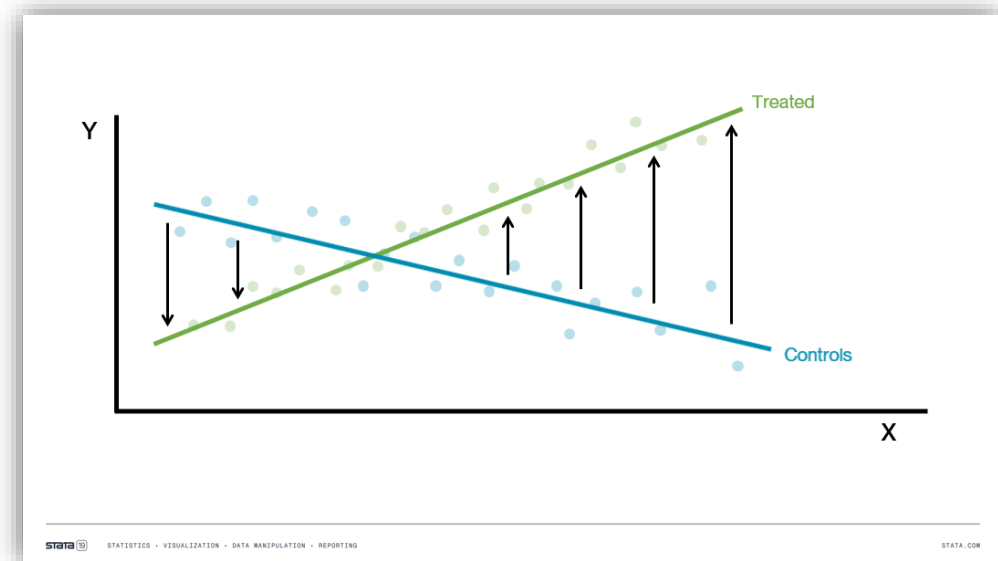
```
cate aipw (ovar catevarlist) (tvar) [if] [in] [, options]
```

ovar is a continuous outcome of interest.

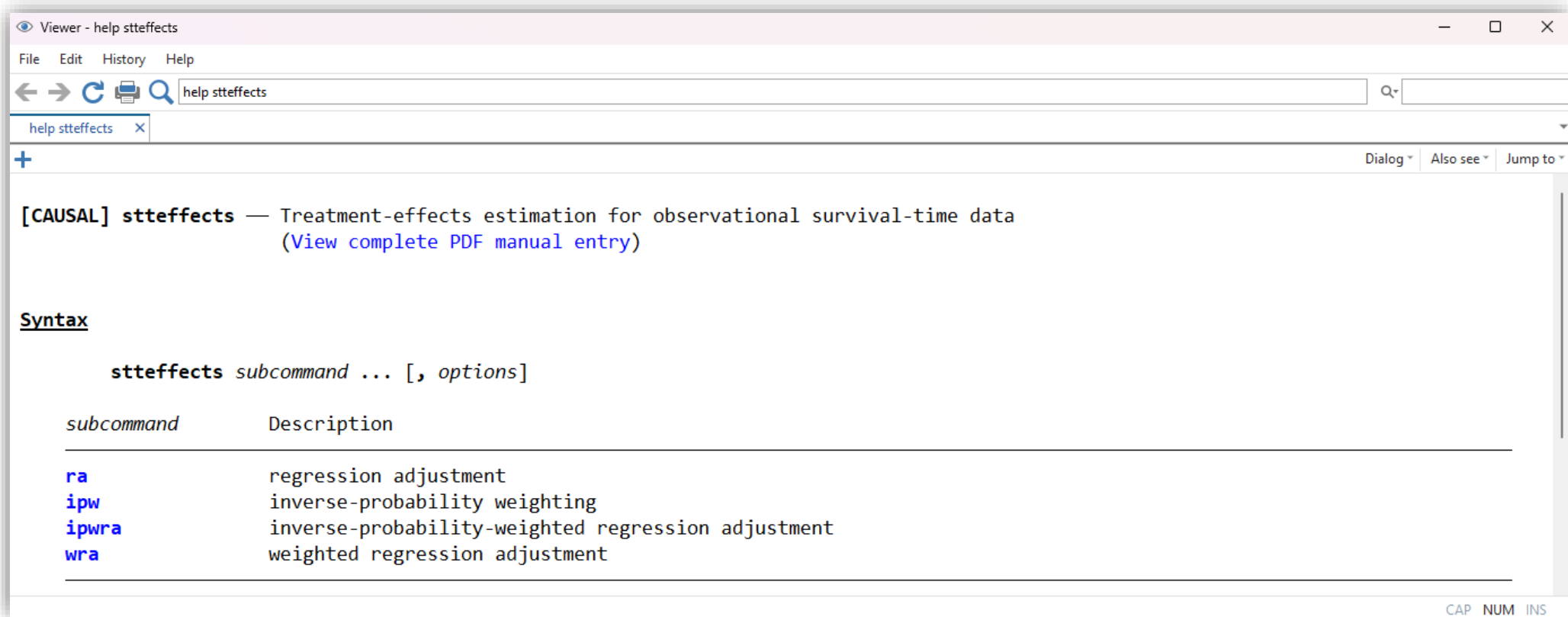
catevarlist specifies the covariates of the CATE model -- the conditioning variables for the treatment effects. *catevarlist* may contain factor variables; see [fvvarlist](#).

tvar must be a binary variable representing the treatment levels.

CAP NUM INS







Viewer - help mediate

File Edit History Help

help mediate

help mediate x

Dialog Also see Jump to

[CAUSAL] mediate — Causal mediation analysis
([View complete PDF manual entry](#))

Syntax

```
mediate (ovar [omvarlist, omodel noconstant]) (mvar [mmvarlist, mmodel noconstant]) (tvar [, continuous(numlist)]) [if] [in]
[weight] [, stat options]
```

ovar is a continuous, binary, or count outcome of interest.
omvarlist specifies the covariates in the outcome model.
mvar is the mediator variable and may be continuous, binary, or count.
mmvarlist specifies the covariates in the mediator model.
tvar is the treatment variable and may be binary, multivalued, or continuous.

<i>omodel</i>	Description
Model	
linear	linear model; the default
<u>expmean</u>	exponential-mean model
logit	logistic regression model
probit	probit regression model
poisson	Poisson model

omodel specifies the model for the outcome variable.

<i>mmodel</i>	Description
Model	
linear	linear model; the default
<u>expmean</u>	exponential-mean model
logit	logistic regression model
probit	probit regression model
poisson	Poisson model

mmodel specifies the model for the mediator variable.

CAP NUM INS

Thank you for joining us!

Installation and use of Stata: tech-support@stata.com
Pricing and upgrades to StataNow: service@stata.com