Causal Mediation Using Stata

Chuck Huber
StataCorp
chuber@stata.com

Oceania Stata Conference
February 6, 2025

Download Website

- You can download all of the slides, datasets and do-files here:

https://tinyurl.com/stata-mediation

Acknowledgement

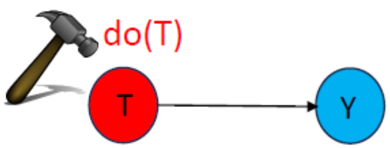
- Aramayis Dallakyan, PhD is a Senior Statistician at StataCorp.
- My slides are heavily inspired by his talk about causal mediation.
- Many of the slides in my talk are literally copied from his talk.
- All errors are my own

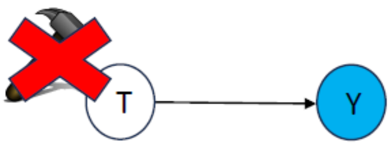
Outline

- Causal Inference
  - Treatment effects with **regress**
  - Treatment effects with **teffects ra**
- Mediation
  - Mediation using **regress**
  - Mediation using **sem** and **estat teffects**
  - Mediation using **sem** and **bootstrap**
- Causal Mediation
  - Continuous outcomes with **mediate**
  - Binary outcomes with **mediate**
  - Count mediators with **mediate**

Causal Inference

- Causal inference tackles the fundamental questions of cause and effect.
- The causal effect aims to compare the outcome when an action T is taken versus the outcome when the action T is withheld.



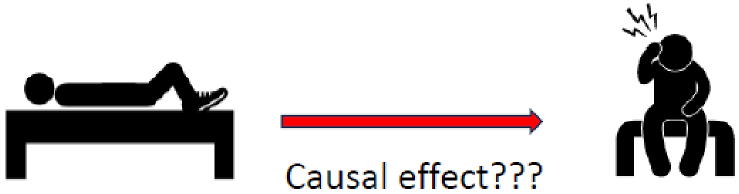


- We refer to action T as an intervention, an exposure, or a treatment.
  - Effect of a treatment/drug/vaccine on a disease;
  - Effect of social media on mental health;
  - Effect of genes on a disease, etc.

Causal Inference

- **Why do we need causality?**
- Why association or statistical dependence is not enough?
- Association does not imply causation!
- The amount of association and the amount of causation can be different

Causal Inference



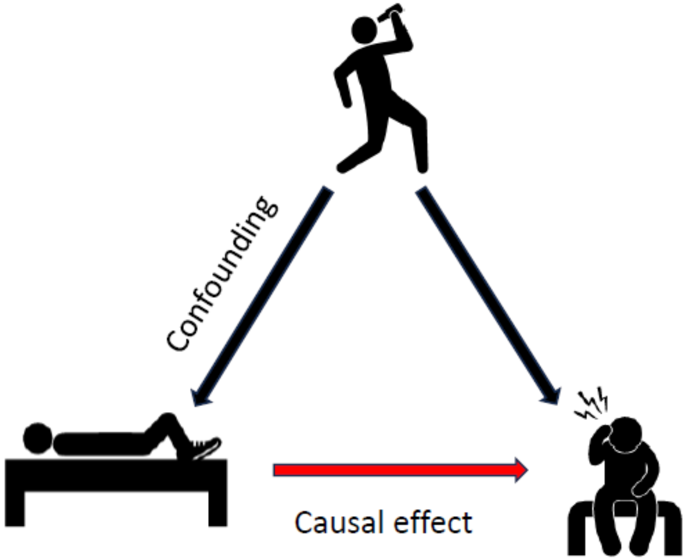
- Suppose we analyze data where the "treatment" is sleeping with shoes on (or not), and the outcome is waking up with a headache (or not) the next day.

- We find that most times when someone wears shoes to bed, that person wakes up with a headache.

- Question: Can we interpret this relationship as causal?

Causal Inference

- One possible explanations for association
  - Both treatment and outcome are caused by a common cause: drinking the night before.
  - Such variables are known as confounders and the association as confounding association.
  - Confounding is the main source of differentiating association from causation.



*Borrowed from Neal (2020)

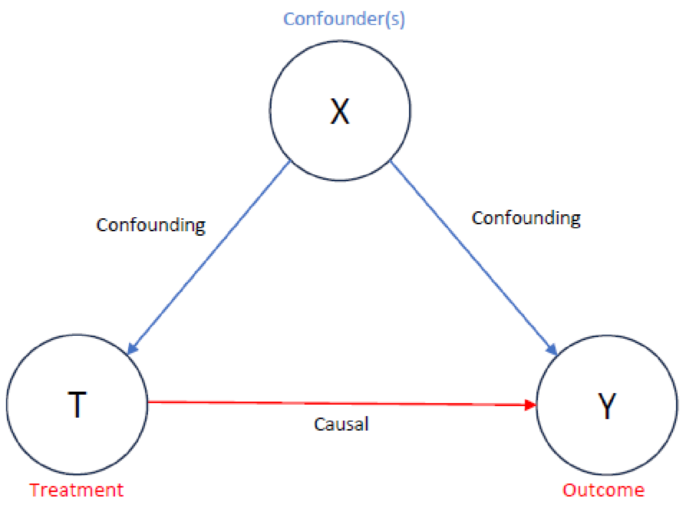
Causal Inference

- Our goal: Learn about causal effects
  - Represent the causal structure
  - Characterize the causal effect

- Notation:
  - **T** ϵ {0,1} denotes treatment assignment: Wearing shoes vs not wearing shoes to bed
  - **Y** denotes the outcome: Headache vs no headache
  - **X** denotes potential confounders that affect both **T** and **Y:** Drinking the previous day

Directed Acyclic Graphs (DAGs)

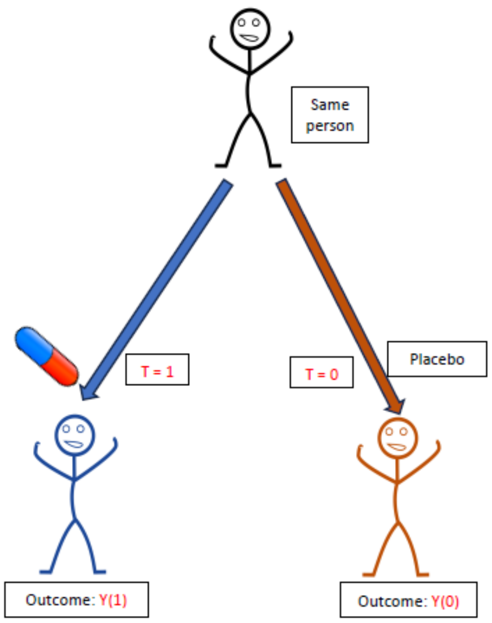
Directed acyclic graphs (DAGs)



- We use DAGs to represent causal relationships and structure.
- Arrows indicate a direct causal effect (not mediated) for at least one subject.
- Informally, **the goal of causal inference** is to estimate the **causal part** of the graph while controlling for the **confounding part**.

Potential Outcomes Framework

- To characterize the causal effect we use the potential outcomes framework.



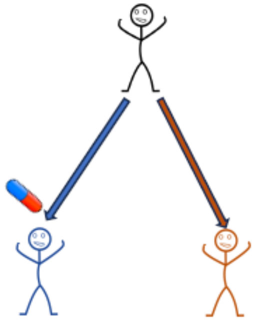
- The potential outcome **Y(T = t) = Y(t)** is the outcome we would have observed had T = t been assigned.
- The causal effect can be measured as Y(1) -Y(0), which is the change due to the treatment keeping everything else the same.

Potential Outcomes Framework

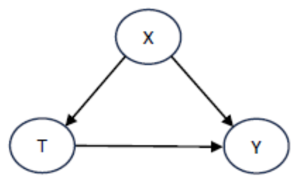
- Fundamental Problem of Causal Inference: Only one of **{Y(1), Y(0)}** is observed.

- The *observed* potential outcome is called **factual.**
- The *unobserved* potential outcome is called **counterfactual**.

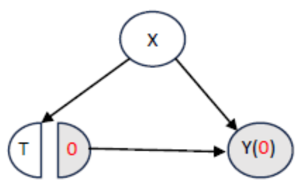
- The causal effect is a **contrast between two parallel worlds**, which we imagine for the same subject.



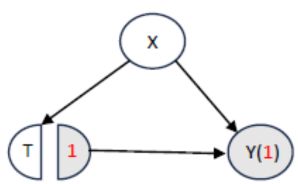
Potential Outcomes Framework



- Actual, observed world.



- In this world everything is the same but **T** is set to 0.

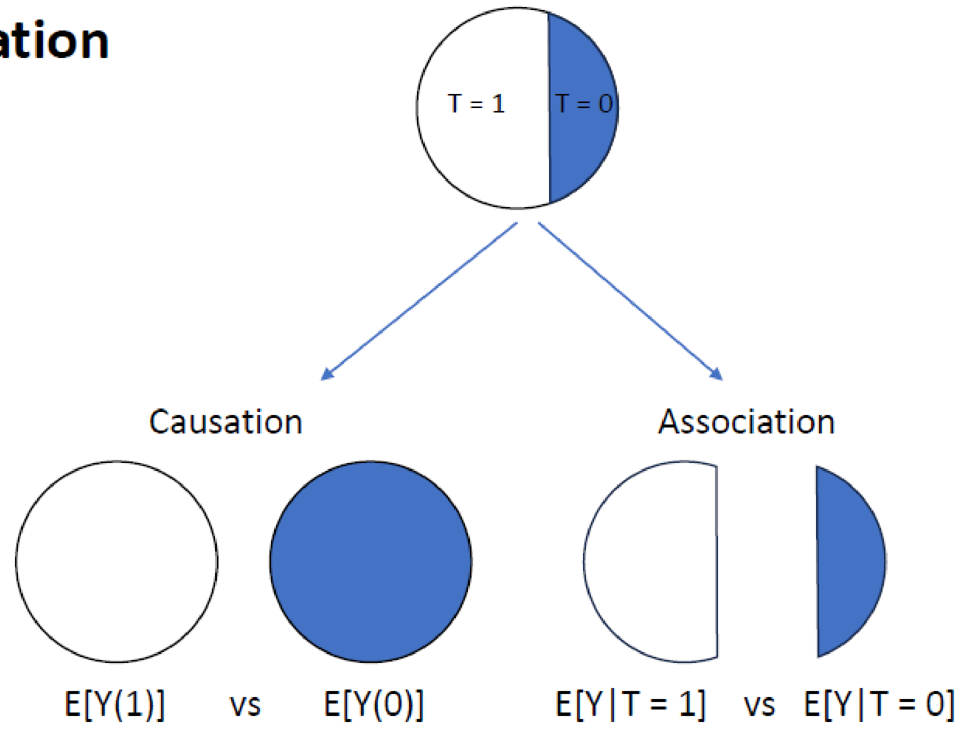


- In this world everything is the same but **T** is set to 1.

- **Note** that compared to the observed world, in imaginary worlds the causal link between **X** and treatment **T** is broken.

Causal Identification

Causal identification



- In general, the causal effect is not the association effect:

$$E[Y(1)] - E[Y(0)] \neq E[Y|T=1] - E[Y|T=0]$$

Outline

- Causal Inference
  - Treatment effects with **`regress`**
  - Treatment effects with **`teffects ra`**
- Mediation
  - Mediation using **`regress`**
  - Mediation using **`sem`** and **`estat teffects`**
  - Mediation using **`sem`** and **`bootstrap`**
- Causal Mediation
  - Continuous outcomes with **`mediate`**
  - Binary outcomes with **`mediate`**
  - Count mediators with **`mediate`**

Treatment Effects With **regress**

```
. use NHANES2015.dta, clear
(Example dataset based on NHANES 2015.  Type 'notes _dta' for more information.)

. describe sbp weight meds_bp

Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
sbp             int     %9.0g               * Systolic Blood Pressure (mmHg)
weight          double  %10.0g              * Weight (kg)
meds_bp         byte    %10.0g     meds     * Taking prescription for hypertension
```

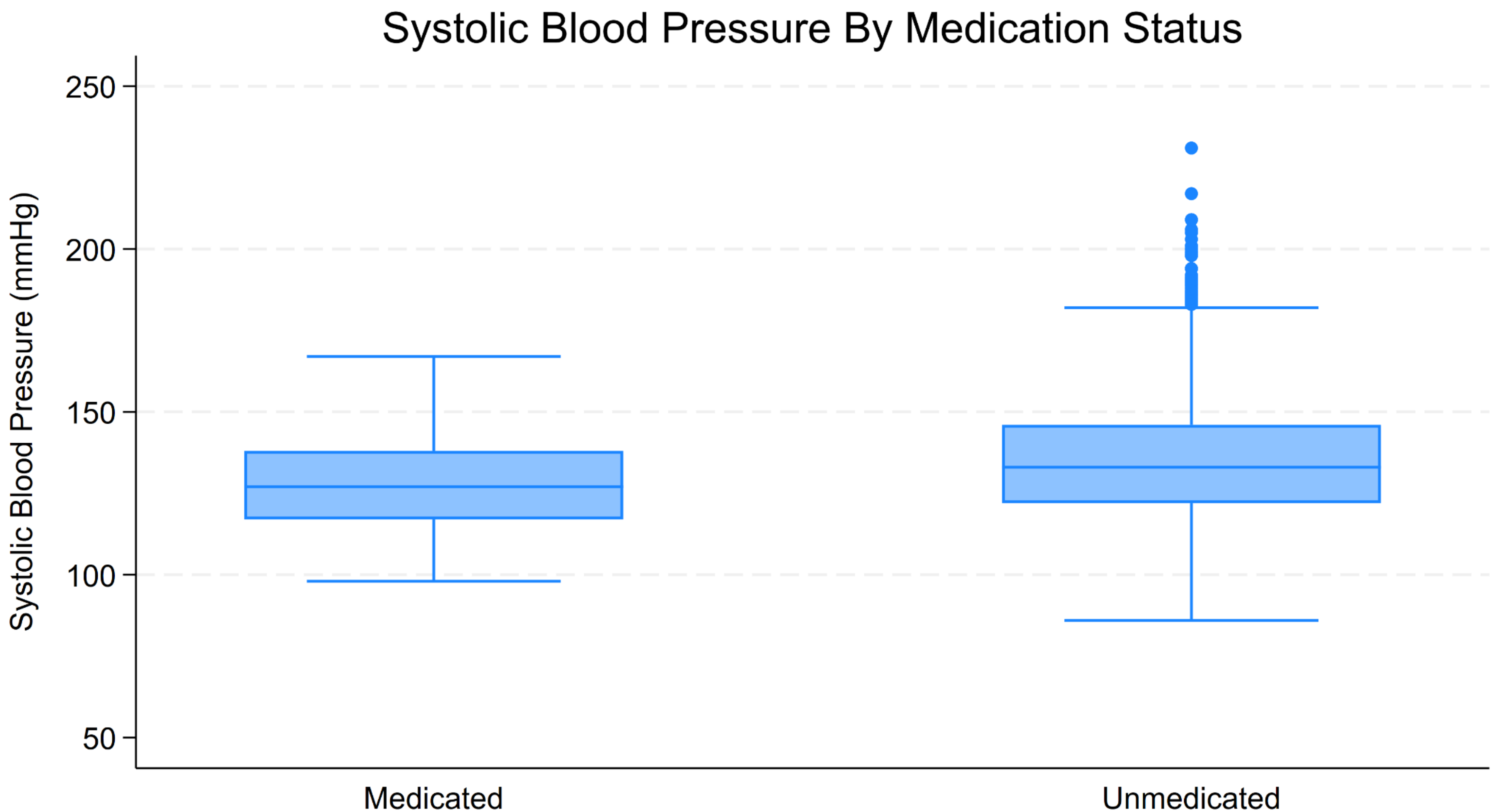
This is observational data, not clinical trial data!

Treatment Effects With **regress**

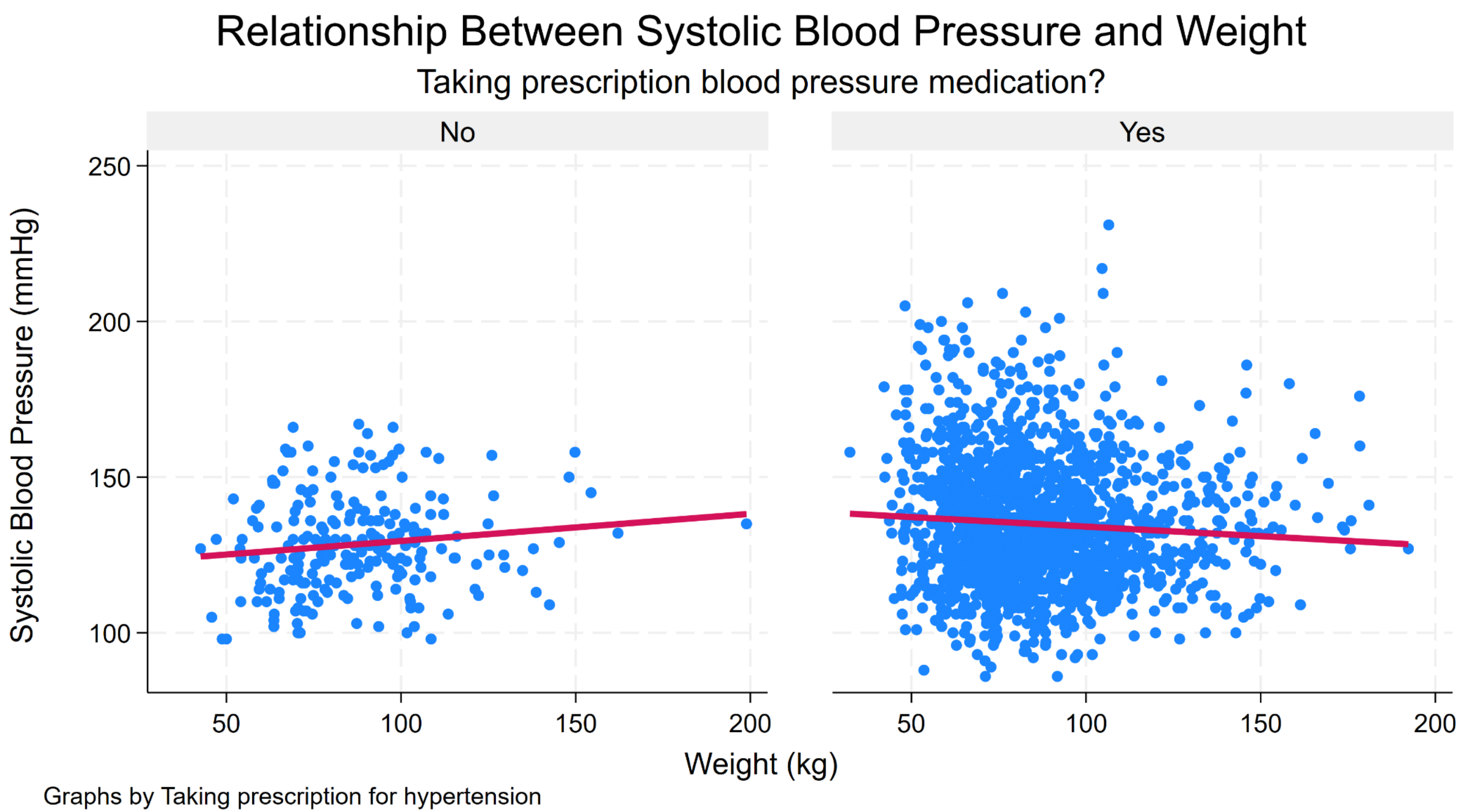
```
. list meds_bp weight sbp in 1/10
```

	meds_bp	weight	sbp
1.	No	53.9	127
2.	No	142.5	109
3.	No	121.2	114
4.	No	96.5	155
5.	No	74.9	112
6.	Yes	74.3	98
7.	Yes	120.7	116
8.	Yes	75.7	130
9.	Yes	62.2	166
10.	Yes	145.1	105

Treatment Effects With `regress`



Treatment Effects With `regress`



Treatment Effects With **regress**

```
. list meds_bp weight sbp obs_sbp0 obs_sbp1 in 1/10
```

	meds_bp	weight	sbp	obs_sbp0	obs_sbp1
1.	No	53.9	127	127	.
2.	No	142.5	109	109	.
3.	No	121.2	114	114	.
4.	No	96.5	155	155	.
5.	No	74.9	112	112	.
6.	Yes	74.3	98	.	98
7.	Yes	120.7	116	.	116
8.	Yes	75.7	130	.	130
9.	Yes	62.2	166	.	166
10.	Yes	145.1	105	.	105

?

Treatment Effects With **`regress`**

```
. regress sbp weight if meds_bp==0

      Source |       SS           df       MS      Number of obs   =       219
-------------+----------------------------------   F(1, 217)       =      3.62
       Model |  872.466244         1  872.466244   Prob > F        =    0.0583
    Residual |  52239.1867       217  240.733579   R-squared       =    0.0164
-------------+----------------------------------   Adj R-squared   =    0.0119
       Total |   53111.653       218  243.631436   Root MSE        =    15.516

------------------------------------------------------------------------------
         sbp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      weight |   .0871461   .0457764     1.90   0.058    -.0030773    .1773695
       _cons |   120.7801   4.152478    29.09   0.000     112.5958    128.9645
------------------------------------------------------------------------------

. generate exp_sbp0 = 120.7801 + .0871461*weight
```

Treatment Effects With **regress**

```
. list meds_bp obs_sbp0 obs_sbp1 exp_sbp0 in 1/10
```

	meds_bp	obs_sbp0	obs_sbp1	exp_sbp0
1.	No	127	.	125.5
2.	No	109	.	133.2
3.	No	114	.	131.3
4.	No	155	.	129.2
5.	No	112	.	127.3
6.	Yes	.	98	127.3
7.	Yes	.	116	131.3
8.	Yes	.	130	127.4
9.	Yes	.	166	126.2
10.	Yes	.	105	133.4

Potential Outcome Y(0)

Treatment Effects With **regress**

```
. regress sbp weight if meds_bp==1

      Source |       SS           df       MS      Number of obs   =     1,715
-------------+----------------------------------   F(1, 1713)      =      8.75
       Model |  3409.90256         1  3409.90256   Prob > F        =    0.0031
    Residual |  667623.832     1,713   389.73954   R-squared       =    0.0051
-------------+----------------------------------   Adj R-squared   =    0.0045
       Total |  671033.734     1,714  391.501595   Root MSE        =    19.742

------------------------------------------------------------------------------
         sbp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      weight |  -.0612475   .0207064    -2.96   0.003    -.1018599    -.020635
       _cons |   140.2318   1.858353    75.46   0.000      136.587    143.8767
------------------------------------------------------------------------------

. gen exp_sbp1 = 140.2318 - .0612475*weight
```

Treatment Effects With **regress**

```
. list meds_bp obs_sbp0 obs_sbp1 exp_sbp0 exp_sbp1 in 1/10
```

	meds_bp	obs_sbp0	obs_sbp1	exp_sbp0	exp_sbp1
1.	No	127	.	125.5	136.9
2.	No	109	.	133.2	131.5
3.	No	114	.	131.3	132.8
4.	No	155	.	129.2	134.3
5.	No	112	.	127.3	135.6
6.	Yes	.	98	127.3	135.7
7.	Yes	.	116	131.3	132.8
8.	Yes	.	130	127.4	135.6
9.	Yes	.	166	126.2	136.4
10.	Yes	.	105	133.4	131.3

Potential Outcome Y(1)

Treatment Effects With **`regress`**

```
. list meds_bp obs_sbp0 obs_sbp1 exp_sbp0 exp_sbp1 in 1/10
```

	meds_bp	obs_sbp0	obs_sbp1	exp_sbp0	exp_sbp1
1.	No	127	.	125.5	136.9
2.	No	109	.	133.2	131.5
3.	No	114	.	131.3	132.8
4.	No	155	.	129.2	134.3
5.	No	112	.	127.3	135.6
6.	Yes	.	98	127.3	135.7
7.	Yes	.	116	131.3	132.8
8.	Yes	.	130	127.4	135.6
9.	Yes	.	166	126.2	136.4
10.	Yes	.	105	133.4	131.3

We have estimated the expected sbp assuming **everyone** took the medication and assuming **no one** took the medication.

Potential Outcome Means (POMs)

The potential-outcome means (POMs) are the means of Y_0 and Y_1 in the population.

```
. mean exp_sbp0 exp_sbp1

Mean estimation                              Number of obs = 1,934

--------------------------------------------------------------
             |       Mean   Std. err.     [95% conf. interval]
-------------+------------------------------------------------
    exp_sbp0 |   128.3497   .0456114      128.2603    128.4392
    exp_sbp1 |   134.9118   .0320563      134.8489    134.9746
--------------------------------------------------------------
```

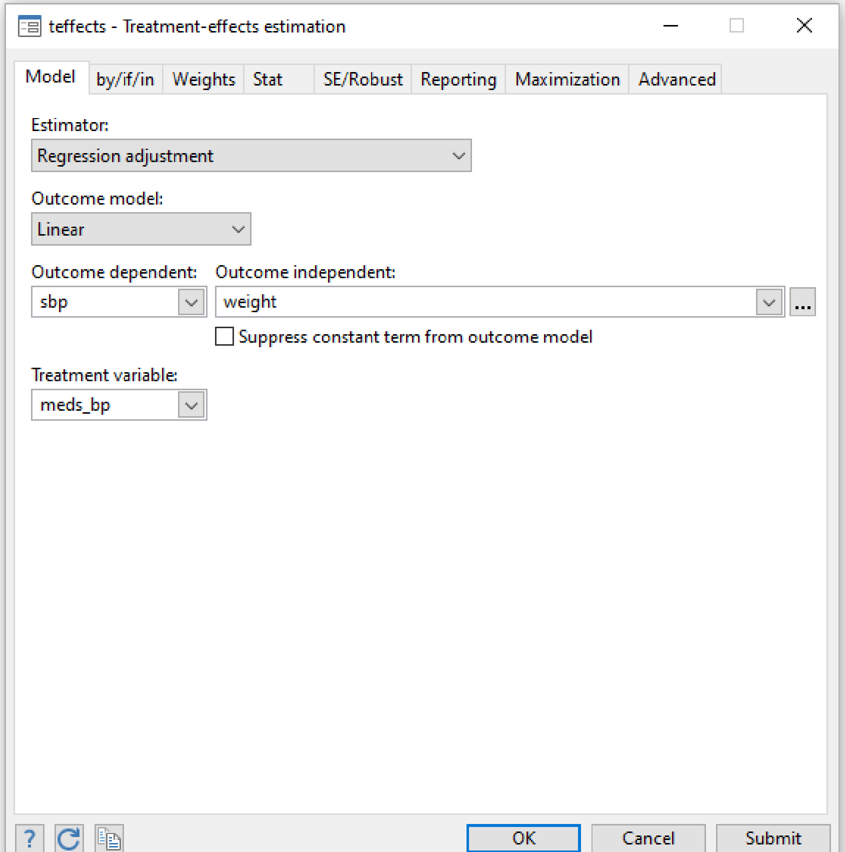
Average Treatment Effect (ATE)

```
. generate te = exp_sbp1 - exp_sbp0

. mean te

Mean estimation                          Number of obs = 1,934
```

	Mean	Std. err.	[95% conf.	interval]
te	6.562066	.0776677	6.409745	6.714387

The average treatment effect (ATE) is the mean of the difference Y(1) – Y(0).

Treatment Effects With **teffects ra**




```
teffects ra (sbp weight) (meds_bp)
```

POMs With **teffects ra**

```
. teffects ra (sbp weight) (meds_bp), pomeans nolog

Treatment-effects estimation                    Number of obs     =      1,934
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
------------------------------------------------------------------------------
             |               Robust
         sbp | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
POmeans      |
     meds_bp |
         No  |   128.3498   1.048327   122.43   0.000     126.2951    130.4044
        Yes  |   134.9118   .4772015   282.71   0.000     133.9765    135.8471
------------------------------------------------------------------------------

. mean exp_sbp0 exp_sbp1

Mean estimation                          Number of obs = 1,934

--------------------------------------------------------------
             |       Mean   Std. err.     [95% conf. interval]
-------------+------------------------------------------------
    exp_sbp0 |   128.3497   .0456114      128.2603    128.4392
    exp_sbp1 |   134.9118   .0320563      134.8489    134.9746
--------------------------------------------------------------
```

ATE With `teffects ra`

```
. teffects ra (sbp weight) (meds_bp), ate nolog

Treatment-effects estimation                    Number of obs     =      1,934
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

sbp	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
meds_bp (Yes vs No)	6.562067	1.153127	5.69	0.000	4.30198	8.822153
POmean						
meds_bp No	128.3498	1.048327	122.43	0.000	126.2951	130.4044

```
. mean te exp_sbp0

Mean estimation                          Number of obs = 1,934
```

	Mean	Std. err.	[95% conf.	interval]
te	6.562066	.0776677	6.409745	6.714387
exp_sbp0	128.3497	.0456114	128.2603	128.4392

Sneak Preview! ATE with **mediate**

```
. mediate (sbp) (weight) (meds_bp), ate nolog

Final EE criterion =  3.83e-29

Causal mediation analysis                                Number of obs = 1,934

Outcome model:      Linear
Mediator model:     Linear
Mediator variable: weight
Treatment type:     Binary
------------------------------------------------------------------------------------
                 |               Robust
             sbp | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
ATE              |
         meds_bp |
    (Yes vs No)  |    6.489727  1.155656     5.62   0.000     4.224683     8.75477
------------------------------------------------------------------------------------
Note: Outcome equation includes treatment-mediator interaction.

. mean te

Mean estimation                          Number of obs = 1,934

--------------------------------------------------------------
             |       Mean   Std. err.     [95% conf. interval]
-------------+------------------------------------------------
          te |   6.562066   .0776677      6.409745    6.714387
--------------------------------------------------------------
```

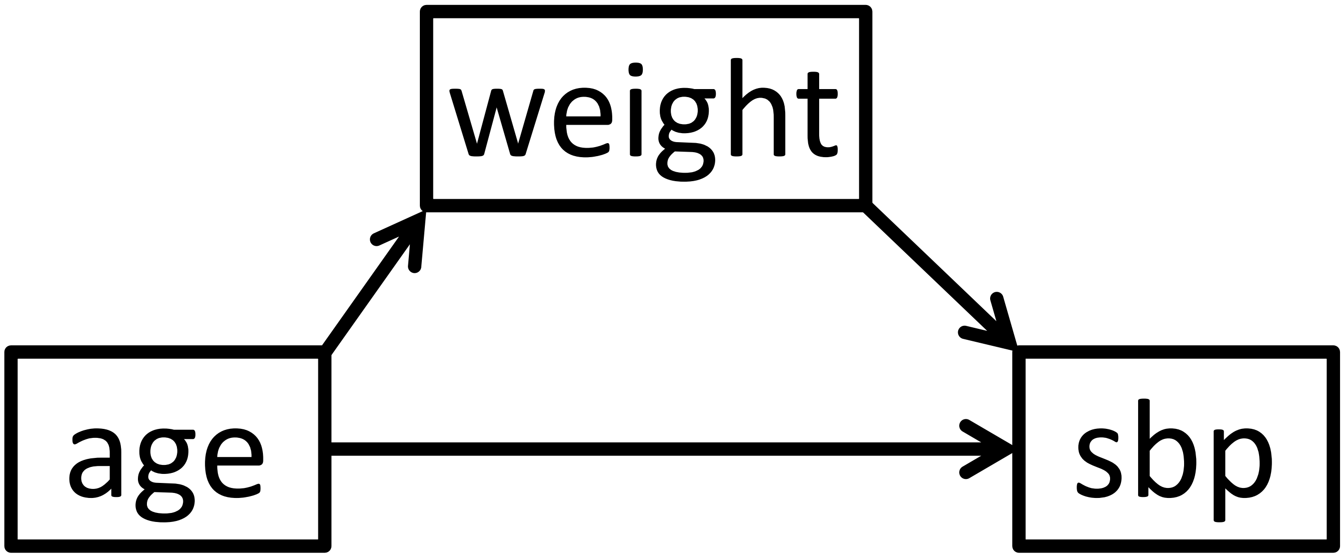
Outline

- Causal Inference
  - ✓ – Treatment effects with **`regress`**
  - ✓ – Treatment effects with **`teffects ra`**
- Mediation
  - Mediation using **`regress`**
  - Mediation using **`sem`** and **`estat teffects`**
  - Mediation using **`sem`** and **`bootstrap`**
- Causal Mediation
  - Continuous outcomes with **`mediate`**
  - Binary outcomes with **`mediate`**
  - Count mediators with **`mediate`**

What is Mediation?

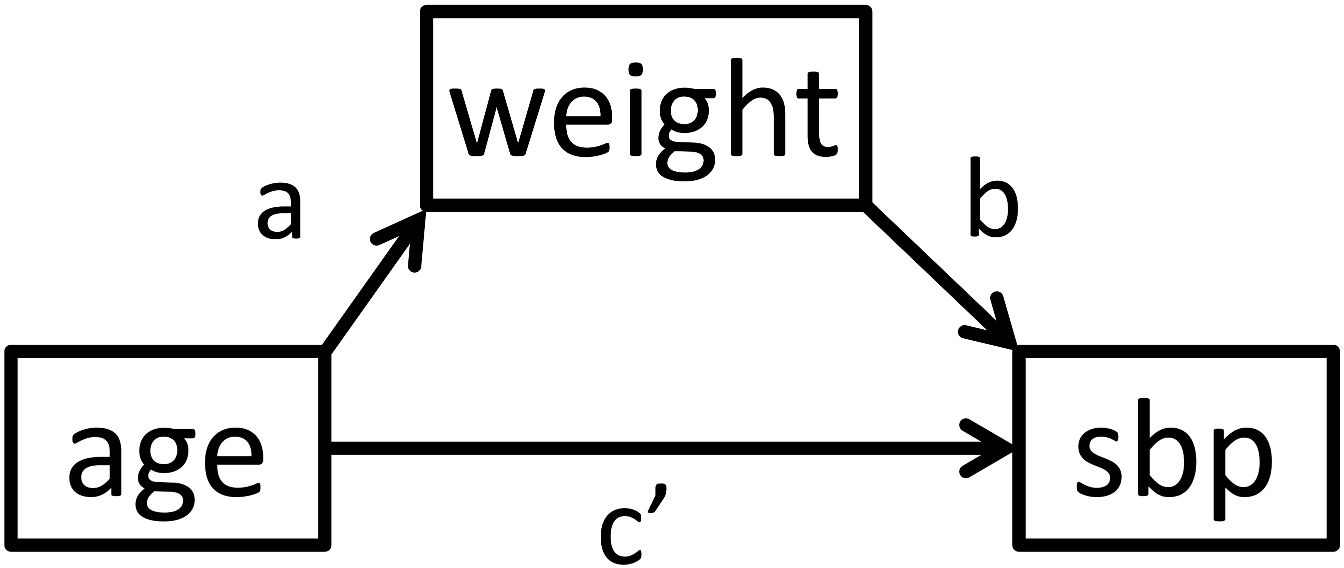
"In general, a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor and the criterion. Mediators explain how external physical events take on internal psychological significance. Whereas moderator variables specify when certain effects will hold, mediators speak to how or why such effects occur." (Baron & Kenny, 1986, pg 1176).

What is Mediation?



Example: Systolic blood pressure (SBP) tends to increase with age. Weight also tends to increase with age. And weight gain is associated with higher SBP. Thus age has a **direct effect** on SBP but age also has an **indirect effect** on SBP through its effect on the mediating variable weight.

What is Mediation?

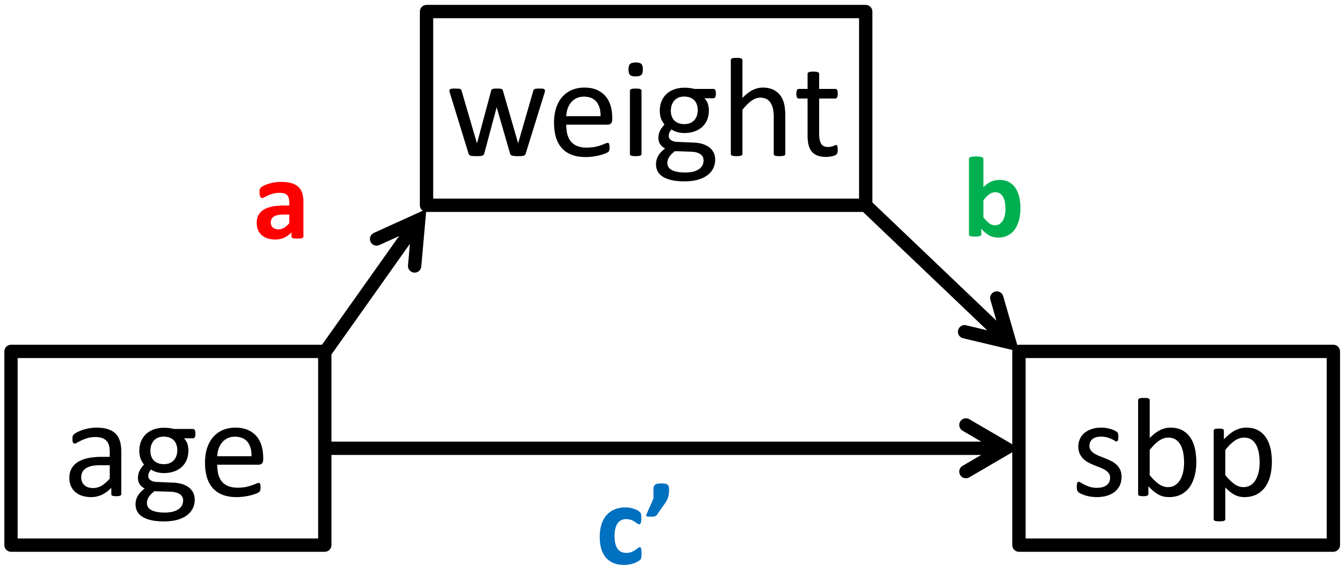


Direct Effect = c’
Indirect Effect = a*b
Total Effect = c’ + a*b

Mediation Using **regress**

```
. use NHANES2015.dta, clear
(Example dataset based on NHANES 2015.  Type 'notes _dta' for more information.)

. describe sbp weight age

Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------------
sbp             int     %9.0g                * Systolic Blood Pressure (mmHg)
weight          double  %10.0g               * Weight (kg)
age             byte    %10.0g               * Age in years at screening
```

Mediation Using **`regress`**



$$weight = \beta_w + \color{red}{a}(age) + e_w$$

$$sbp = \beta_s + \color{blue}{c'}(age) + \color{green}{b}(weight) + e_s$$

Estimate a

```
. regress weight age if !missing(sbp)

      Source |       SS           df       MS      Number of obs   =     7,299
-------------+----------------------------------   F(1, 7297)      =   1054.97
       Model |  567192.719         1  567192.719   Prob > F        =    0.0000
    Residual |  3923147.98     7,297  537.638479   R-squared       =    0.1263
-------------+----------------------------------   Adj R-squared   =    0.1262
       Total |   4490340.7     7,298  615.283735   Root MSE        =    23.187

------------------------------------------------------------------------------
      weight | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |    .395699   .0121827    32.48   0.000     .3718173    .4195807
       _cons |   58.89463   .5508451   106.92   0.000     57.81481    59.97444
------------------------------------------------------------------------------
```

```
. scalar a = 0.395699
```

a = 0.395699

Estimate b and c′

```
. regress sbp age weight

      Source |       SS           df       MS      Number of obs   =     7,299
-------------+----------------------------------   F(2, 7296)      =   2183.07
       Model |  915467.252         2  457733.626   Prob > F        =    0.0000
    Residual |  1529784.89     7,296  209.674464   R-squared       =    0.3744
-------------+----------------------------------   Adj R-squared   =    0.3742
       Total |  2445252.14     7,298  335.057844   Root MSE        =     14.48

------------------------------------------------------------------------------
         sbp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .4293534   .0081394    52.75   0.000     .4133978    .4453091
      weight |   .1348724   .0073106    18.45   0.000     .1205414    .1492034
       _cons |   93.07427   .5511034   168.89   0.000     91.99395    94.15459
------------------------------------------------------------------------------
```

```
. scalar cprime = 0.4293534
```

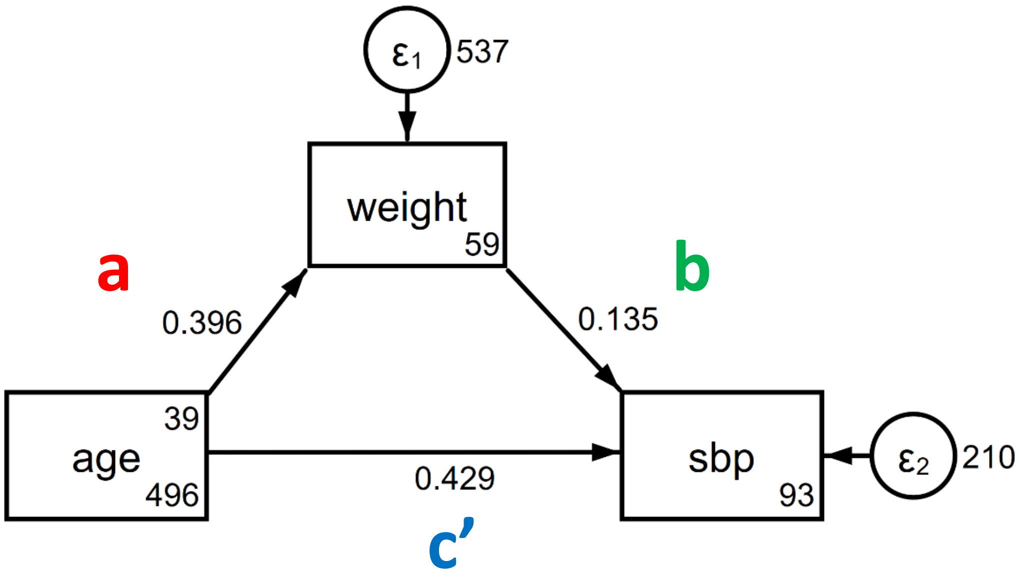
```
. scalar b = 0.1348724
```

$c' = 0.4293534$

$b = 0.1348724$

Direct, Indirect and Total Effects

```
. // DIRECT EFFECT OF age ON sbp
. display "The Direct Effect of age on sbp is " cprime
The Direct Effect of age on sbp is .4293534

.
. // INDIRECT EFFECT OF age ON sbp
. scalar IE = scalar(a)*scalar(b)

. display "The Indirect effect of age on sbp is " IE
The Indirect effect of age on sbp is .05336887

.
. // TOTAL EFFECT OF age ON sbp
. scalar c = cprime + scalar(a)*scalar(b)

. display "The Total Effect of age on sbp is " c
The Total Effect of age on sbp is .48272227
```



The **direct effect** of age on SBP is the estimated coefficient for path c' which is 0.429. Each additional year of age is associated with a 0.429 mm/Hg increase of SBP.

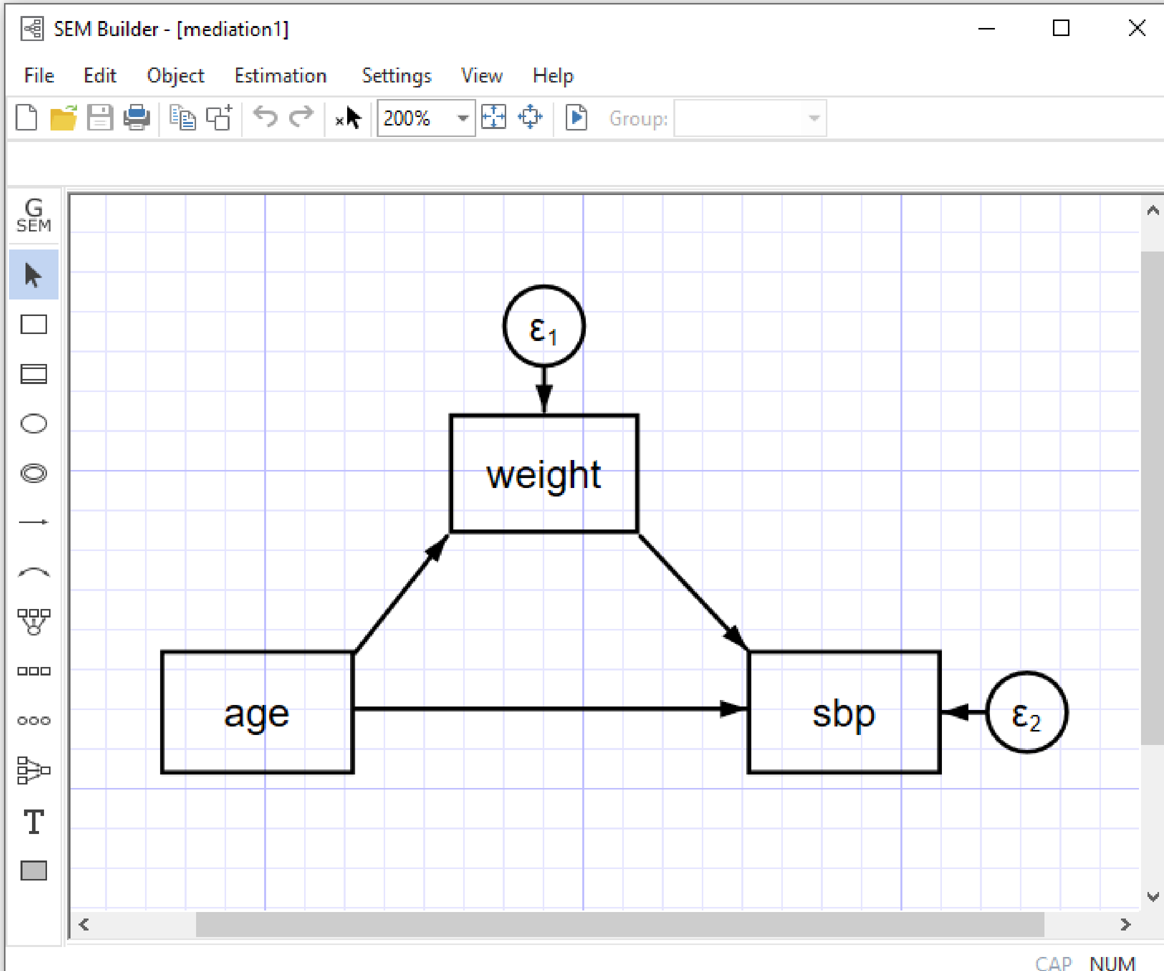
The **indirect effect** of age on SBP through weight is more subtle. Each additional year of age is associated with a 0.396 kg increase in weight. And each additional kilogram increase of weight is associated with a 0.135 mm/Hg increase in SBP. So the estimate of the indirect effect is the product of the coefficients a and b which equals 0.053.

The **total effect** of age on SBP is the sum of the direct and indirect effects. Thus each year of age is associated with a 0.483 mm/Hg increase in SBP.

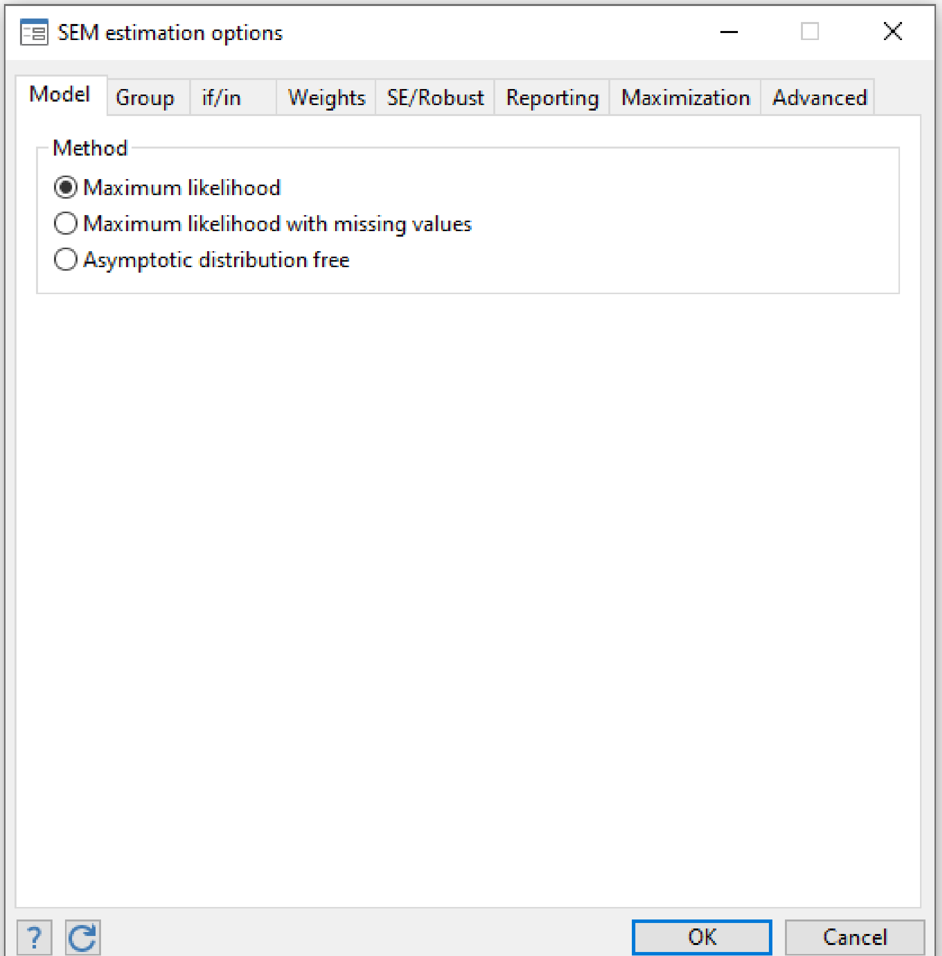
Outline

- Causal Inference
  - ✓ – Treatment effects with **`regress`**
  - ✓ – Treatment effects with **`teffects ra`**
- Mediation
  - ✓ – Mediation using **`regress`**
  - Mediation using **`sem`** and **`estat teffects`**
  - Mediation using **`sem`** and **`bootstrap`**
- Causal Mediation
  - Continuous outcomes with **`mediate`**
  - Binary outcomes with **`mediate`**
  - Count mediators with **`mediate`**

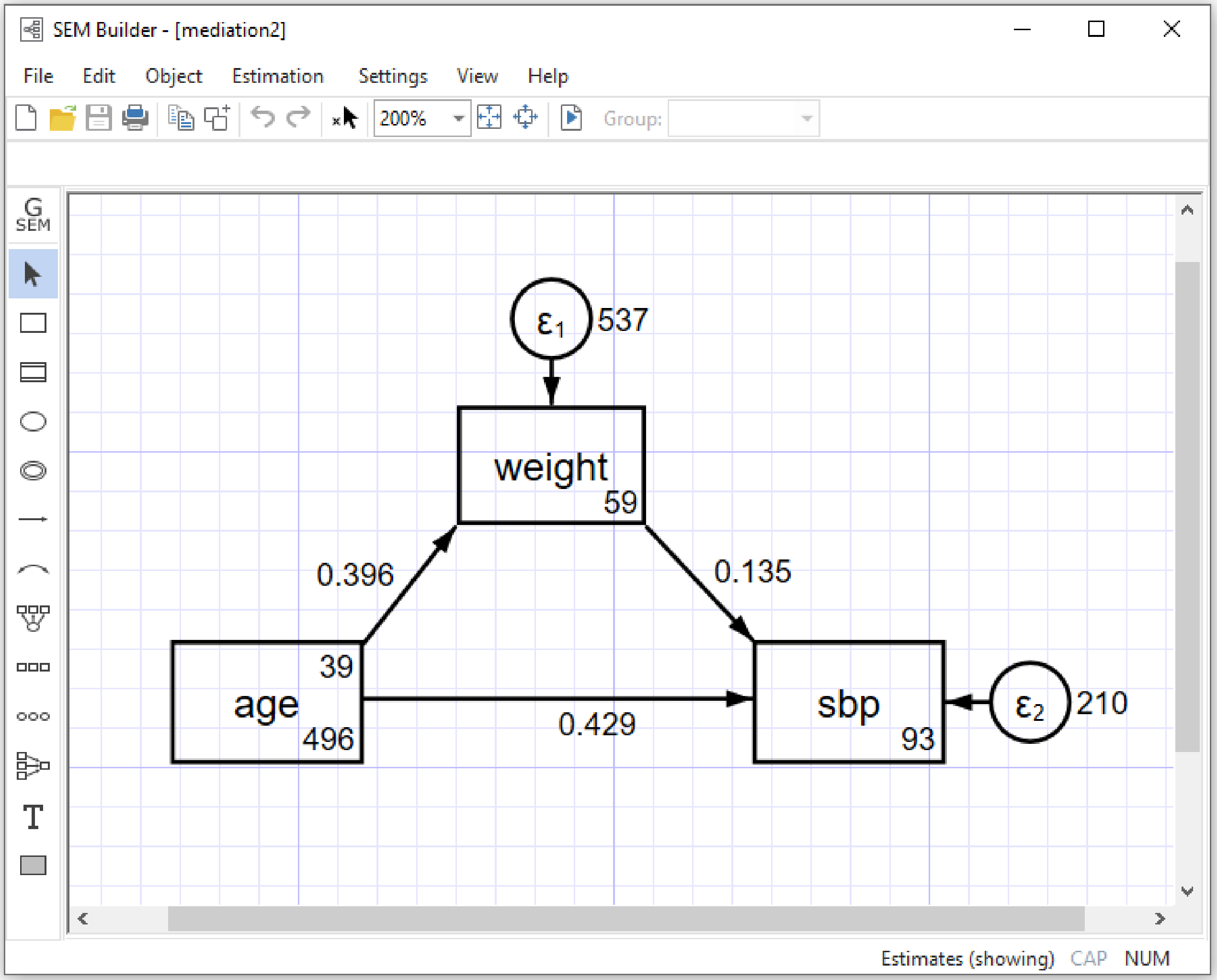
Mediation Using **sem**



Mediation Using **sem**



Mediation Using **sem**



SEM Parameter Estimates

sem (sbp <- age weight) (weight <- age)

```
--------------------------------------------------------------------------------
               |                 OIM
               | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
---------------+----------------------------------------------------------------
Structural     |
  sbp          |
        weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
           age |   .4293534   .0081378    52.76   0.000     .4134037    .4453032
         _cons |   93.07427   .5509902   168.92   0.000     91.99435    94.15419
  -------------+----------------------------------------------------------------
  weight       |
           age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
         _cons |   58.89463   .5507696   106.93   0.000     57.81514    59.97412
---------------+----------------------------------------------------------------
    var(e.sbp) |   209.5883   3.469368                      202.8976    216.4996
 var(e.weight) |   537.4912   8.897226                      520.3328    555.2154
--------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(0) = 0.00                 Prob > chi2 =   .
```

Mediation Using **sem** and **estat teffects**

```
. estat teffects

Direct effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4293534   .0081378    52.76   0.000     .4134037    .4453032
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------

Indirect effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |          0  (no path)
         age |   .0533689   .0033263    16.04   0.000     .0468495    .0598882
  -----------+----------------------------------------------------------------
  weight     |
         age |          0  (no path)
------------------------------------------------------------------------------

Total effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4827223   .0077819    62.03   0.000     .4674701    .4979745
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------
```

$\Big\}$ Direct effects

$\Big\}$ Indirect effects

$\Big\}$ Total effects

Mediation Using **sem** and **estat teffects**

```
. estat teffects
```

```
. estat teffects

Direct effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4293534   .0081378    52.76   0.000     .4134037    .4453032
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------

Indirect effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |          0  (no path)
         age |   .0533689   .0033263    16.04   0.000     .0468495    .0598882
  -----------+----------------------------------------------------------------
  weight     |
         age |          0  (no path)
------------------------------------------------------------------------------

Total effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4827223   .0077819    62.03   0.000     .4674701    .4979745
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------
```

```
Direct effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4293534   .0081378    52.76   0.000     .4134037    .4453032
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------
```

Mediation Using **sem** and **estat teffects**

```
. estat teffects
```

```
. estat teffects

Direct effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4293534   .0081378    52.76   0.000     .4134037    .4453032
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------

Indirect effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |          0  (no path)
         age |   .0533689   .0033263    16.04   0.000     .0468495    .0598882
  -----------+----------------------------------------------------------------
  weight     |
         age |          0  (no path)
------------------------------------------------------------------------------

Total effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4827223   .0077819    62.03   0.000     .4674701    .4979745
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------
```

```
Indirect effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |          0  (no path)
         age |   .0533689   .0033263    16.04   0.000     .0468495    .0598882
  -----------+----------------------------------------------------------------
  weight     |
         age |          0  (no path)
------------------------------------------------------------------------------
```

Mediation Using **sem** and **estat teffects**

. estat teffects

```
. estat teffects

Direct effects

------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4293534   .0081378    52.76   0.000     .4134037    .4453032
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------

Indirect effects

------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |          0  (no path)
         age |   .0533689   .0033263    16.04   0.000     .0468495    .0598882
  -----------+----------------------------------------------------------------
  weight     |
         age |          0  (no path)
------------------------------------------------------------------------------

Total effects

------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4827223   .0077819    62.03   0.000     .4674701    .4979745
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------
```

```
Total effects

------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  sbp        |
      weight |   .1348724   .0073091    18.45   0.000     .1205467     .149198
         age |   .4827223   .0077819    62.03   0.000     .4674701    .4979745
  -----------+----------------------------------------------------------------
  weight     |
         age |    .395699   .0121811    32.48   0.000     .3718246    .4195735
------------------------------------------------------------------------------
```

Outline

- Causal Inference
  - ✓ – Treatment effects with **`regress`**
  - ✓ – Treatment effects with **`teffects ra`**
- Mediation
  - ✓ – Mediation using **`regress`**
  - ✓ – Mediation using **`sem`** and **`estat teffects`**
  - – Mediation using **`sem`** and **`bootstrap`**
- Causal Mediation
  - – Continuous outcomes with **`mediate`**
  - – Binary outcomes with **`mediate`**
  - – Count mediators with **`mediate`**

The **`coeflegend`** option

```
sem (sbp <- age weight) (weight <- age), coeflegend
```

	Coefficient	Legend
Structural		
sbp		
weight	.1348724	_b[sbp:weight]
age	.4293534	_b[sbp:age]
_cons	93.07427	_b[sbp:_cons]
weight		
age	.395699	_b[weight:age]
_cons	58.89463	_b[weight:_cons]
var(e.sbp)	209.5883	_b[/var(e.sbp)]
var(e.weight)	537.4912	_b[/var(e.weight)]

The `bootstrap` prefix

```
bootstrap direct   = (_b[sbp:age])                               ///
          indirect = (_b[weight:age] * _b[sbp:weight])           ///
          total    = (_b[sbp:age] + _b[weight:age] * _b[sbp:weight]) ///
          , nodots : sem (sbp <- age weight) (weight <- age)
```

The **bootstrap** prefix

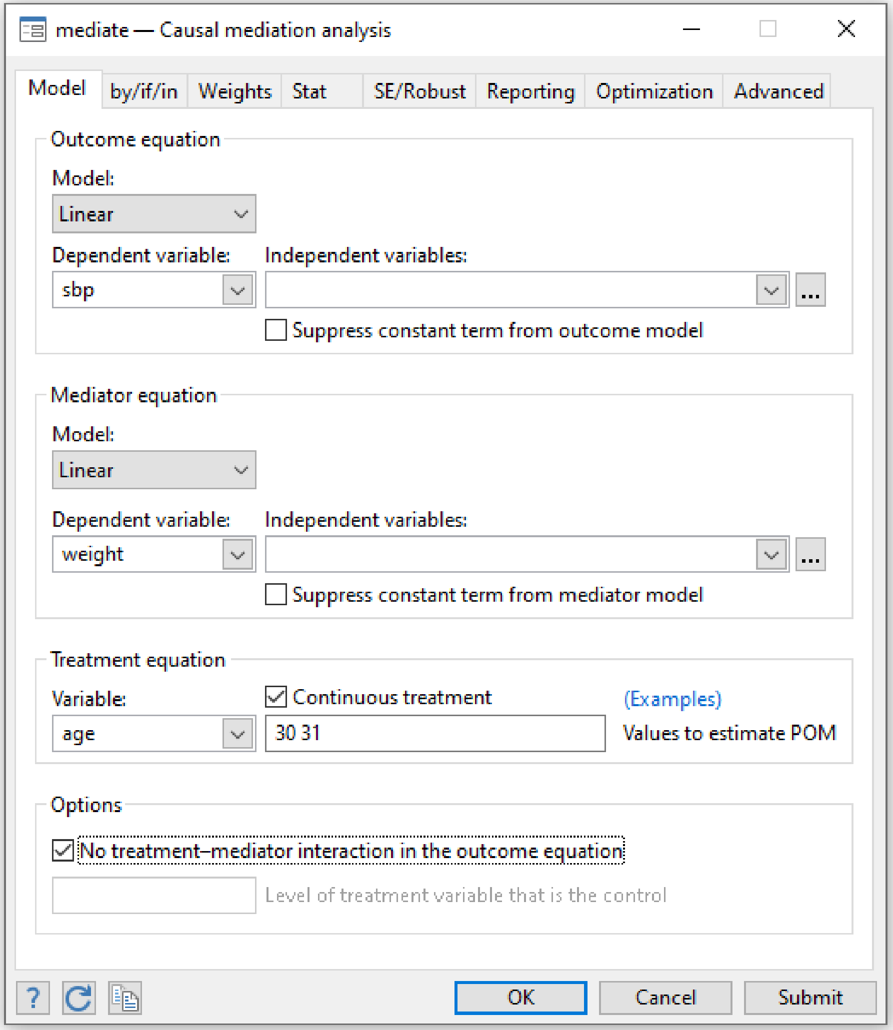
```
. bootstrap direct   = (_b[sbp:age])                                     ///
>           indirect = (_b[weight:age] * _b[sbp:weight])                ///
>           total    = (_b[sbp:age] + _b[weight:age] * _b[sbp:weight])  ///
>           , nodots reps(500) : sem (sbp <- age weight) (weight <- age)

Bootstrap results                                        Number of obs = 7,299
                                                         Replications  =   500

      Command: sem (sbp <- age weight) (weight <- age)
       direct: _b[sbp:age]
     indirect: _b[weight:age] * _b[sbp:weight]
        total: _b[sbp:age] + _b[weight:age] * _b[sbp:weight]
```

	Observed coefficient	Bootstrap std. err.	z	P>|z|	Normal-based [95% conf.	interval]
direct	.4293534	.0099233	43.27	0.000	.4099041	.4488027
indirect	.0533689	.0034491	15.47	0.000	.0466088	.060129
total	.4827223	.0085877	56.21	0.000	.4658908	.4995538

Outline

- Causal Inference
  - ✓ – Treatment effects with **`regress`**
  - ✓ – Treatment effects with **`teffects ra`**
- Mediation
  - ✓ – Mediation using **`regress`**
  - ✓ – Mediation using **`sem`** and **`estat teffects`**
  - ✓ – Mediation using **`sem`** and **`bootstrap`**
- Causal Mediation
  - – Continuous outcomes with **`mediate`**
  - – Binary outcomes with **`mediate`**
  - – Count mediators with **`mediate`**

Continuous Outcomes With `mediate`



Continuous Outcomes With **mediate**

```
. mediate (sbp) (weight) (age, continuous(30 31)), ///
>         adte aite ate nolog nointeract nolegend

Final EE criterion =  3.34e-28

Causal mediation analysis                           Number of obs = 7,299

Outcome model:      Linear
Mediator model:     Linear
Mediator variable: weight
Treatment type:     Continuous
------------------------------------------------------------------------------
             |               Robust
         sbp | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
AITE         |
         age |
   (1 vs 0)  |   .0533689   .0034635    15.41   0.000     .0465805    .0601573
-------------+----------------------------------------------------------------
ADTE         |
         age |
   (1 vs 0)  |   .4293534   .0094489    45.44   0.000     .4108339     .447873
-------------+----------------------------------------------------------------
ATE          |
         age |
   (1 vs 0)  |   .4827223   .0082838    58.27   0.000     .4664864    .4989582
------------------------------------------------------------------------------
Note: Outcome equation does not include treatment-mediator interaction.
```

Direct effect

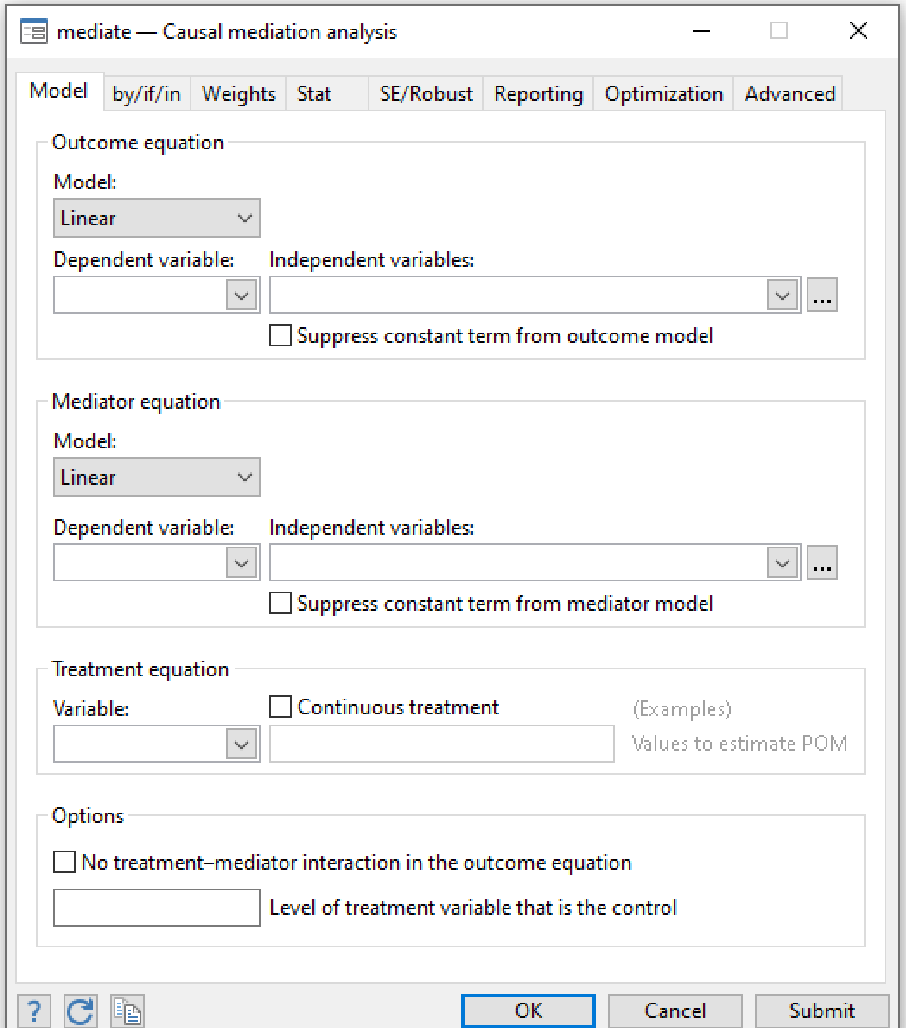
Indirect effect

Total effect

Auxiliary Equations With **aequations**

```
. mediate (sbp) (weight) (age, continuous(30 31)), ///
>         ate nolog nolegend aequations

Final EE criterion =  8.07e-29

Causal mediation analysis                               Number of obs = 7,299

Outcome model:      Linear
Mediator model:     Linear
Mediator variable:  weight
Treatment type:     Continuous
───────────────┬────────────────────────────────────────────────────────────────
               │              Robust
           sbp │ Coefficient  std. err.      z    P>|z|     [95% conf. interval]
───────────────┼────────────────────────────────────────────────────────────────
ATE            │
           age │
     (1 vs 0)  │   .4846903    .008329    58.19   0.000     .4683657    .5010148
───────────────┼────────────────────────────────────────────────────────────────
sbp            │
           age │     .67264   .0288097    23.35   0.000     .6161741    .7291059
        weight │   .2547316   .0115282    22.10   0.000     .2321367    .2773265
               │
c.weight#c.age │  -.0034775   .0003704    -9.39   0.000    -.0042035   -.0027515
               │
         _cons │   85.44831   .7316597   116.79   0.000     84.01428    86.88233
───────────────┼────────────────────────────────────────────────────────────────
weight         │
           age │    .395699   .0118989    33.25   0.000     .3723775    .4190205
         _cons │   58.89463   .5597933   105.21   0.000     57.79745     59.9918
───────────────┴────────────────────────────────────────────────────────────────
Note: Outcome equation includes treatment–mediator interaction.
```

Continuous Outcomes With `mediate`



Continuous Outcomes With `mediate`

```
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.

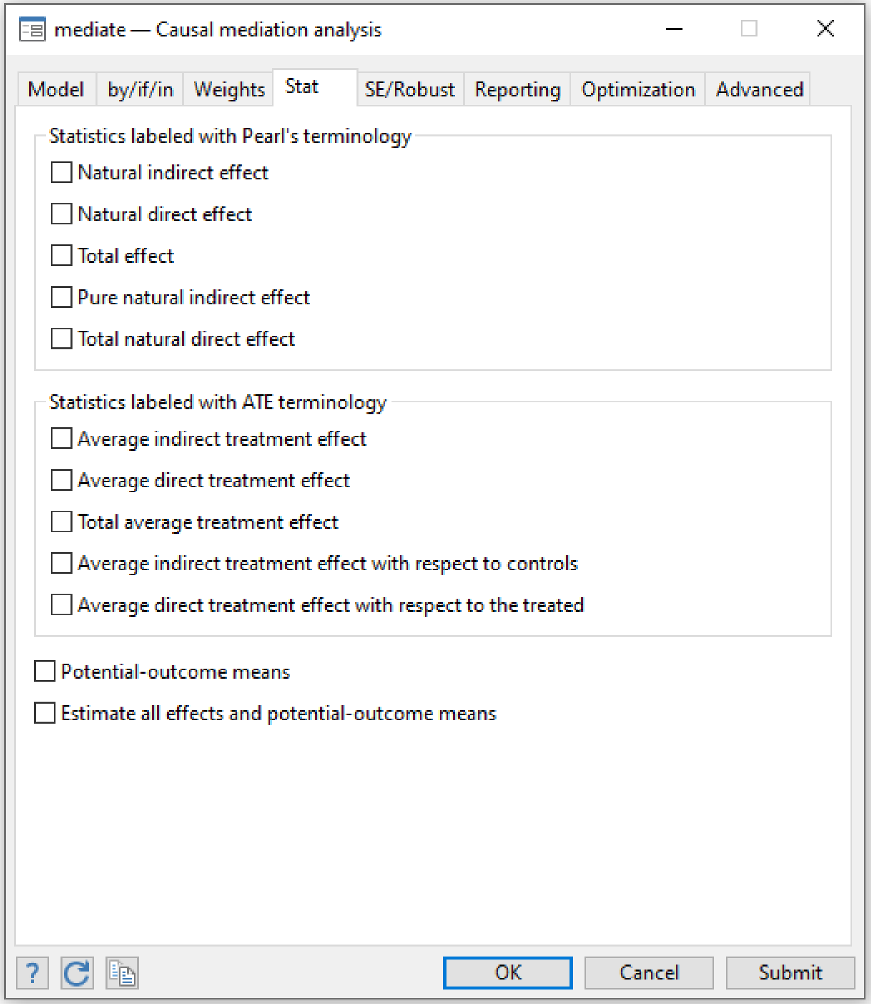
tvar is the treatment variable and may be binary, multivalued, or continuous.

Continuous Outcomes With **`mediate`**

Outcome \ Mediator	linear	logit	probit	Poisson	exp. mean
linear	X	X	X	X	X
logit		X	X	X	
probit	X	X	X	X	X
Poisson	X	X	X	X	X
exp. mean	X	X	X	X	X

Note: X indicates a supported model combination

Continuous Outcomes With **`mediate`**



Continuous Outcomes With **mediate**

Stat

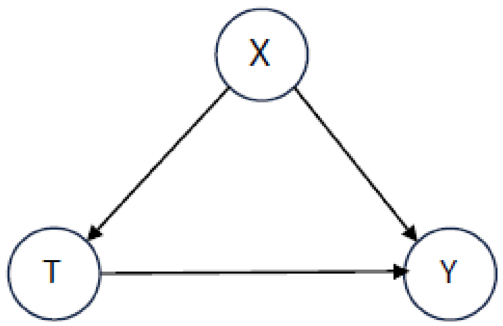
stat specifies the statistics to be estimated. You may select from among five effects, each of which can be labeled according to terminology used by Pearl and others or by ATE terminology. In addition to effects, you may request that potential-outcome means be reported. The default is `nie nde te`.

stat may be one or more of the following:

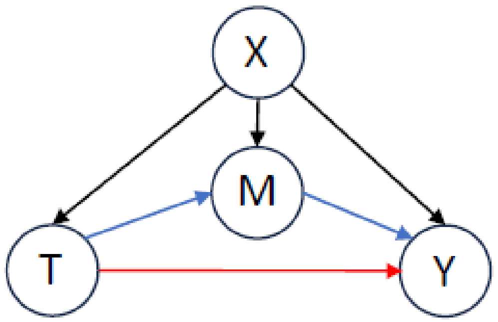
stat	Definition
`nie`	natural indirect effect
`nde`	natural direct effect
`te`	total effect
`pnie`	pure natural indirect effect
`tnde`	total natural direct effect
`aite`	average indirect treatment effect; synonym for `nie`
`adte`	average direct treatment effect; synonym for `nde`
`ate`	average treatment effect; synonym for `te`
`aitec`	average indirect treatment effect with respect to controls; synonym for `pnie`
`adtet`	average direct treatment effect with respect to the treated; synonym for `tnde`
`pomeans`	potential-outcome means

`all` specifies that all effects and potential-outcome means be estimated; specifying `all` is equivalent to specifying `nie nde te pnie tnde pomeans`. When option `ateterms` is specified, `all` is equivalent to specifying `aite adte ate aitec adtet pomeans`.

Causal Mediation



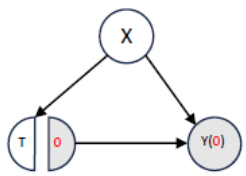
Suppose using the Fundamental Steps of Causal Analysis (FSCA), a researcher concluded that age has a causal effect on SBP.

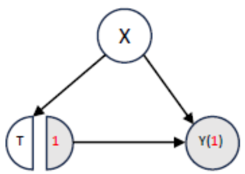


But the researcher wonders whether the increase in SBP is a consequence of the effect of age (T) on weight (M) which increases SBP (Y)

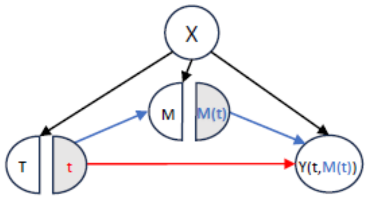
That is, the researcher is interested in decomposing the total effect of T on Y into the indirect causal pathway mediated by M and the direct pathway not mediated by M.

Preparing For Causal Identification





- Recall that our interest is in the contrast **Y(1) - Y(0)**
- For mediation, the idea is to split the contrast **Y(1) - Y(0)** into two other contrasts using a third potential outcome **M(t).**



- We introduce a new type of outcome **Y(t,m)**, which corresponds to the potential outcome when we set **T = t** and **M=m**.

- Note the familiar **Y(1) = Y[1, M(1)]** and **Y(0) = Y[0,M(0)]**.

Potential Outcome Means With **mediate**

```
. mediate (sbp) (weight) (age, continuous(30 31)), pomeans nolegend

Iteration 0:   EE criterion =  3.203e-25
Iteration 1:   EE criterion =  2.057e-28

Causal mediation analysis                              Number of obs = 7,299

Outcome model:      Linear
Mediator model:     Linear
Mediator variable:  weight
Treatment type:     Continuous
```

sbp	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
Y0M0	116.2711	.1742669	667.20	0.000	115.9295	116.6126
Y1M0	116.6976	.1751862	666.13	0.000	116.3543	117.041
Y0M1	116.3306	.1740157	668.51	0.000	115.9895	116.6716
Y1M1	116.7558	.1747872	667.99	0.000	116.4132	117.0983

Note: Outcome equation includes treatment-mediator interaction.

POMs With **gsem** and **nlcom**

```
. gsem (sbp <- c.age##c.weight) (weight <- age), ///
>              nolog noheader listwise coeflegend
```

	Coefficient	Legend
sbp		
weight	.2547316	_b[sbp:weight]
age	.67264	_b[sbp:age]
c.age#c.weight	-.0034775	_b[sbp:c.age#c.weight]
_cons	85.44831	_b[sbp:_cons]
weight		
age	.395699	_b[weight:age]
_cons	58.89463	_b[weight:_cons]
var(e.sbp)	206.705	_b[/var(e.sbp)]
var(e.weight)	537.4912	_b[/var(e.weight)]

POMs With gsem and nlcom

Expected value of SBP assuming everyone is 30 years old (not treated)

```
nlcom                                                                         ///
(Y0M0:   _b[sbp:_cons]                                                        ///
       + _b[sbp:age]*30                                                       ///
       + _b[sbp:weight]          * (_b[weight:_cons] + _b[weight:age]*30)     ///
       + _b[sbp:c.weight#c.age] * (_b[weight:_cons] + _b[weight:age]*30) * 30)
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
Y0M0	116.2711	.2027684	573.42	0.000	115.8737	116.6685

POMs With **gsem** and **nlcom**

Expected value of SBP assuming everyone is 31 years old (treated)

```
nlcom                                                                          ///
(Y1M1:  _b[sbp:_cons]                                                          ///
      + _b[sbp:age]*31                                                         ///
      + _b[sbp:weight]         * (_b[weight:_cons] + _b[weight:age]*31)        ///
      + _b[sbp:c.weight#c.age] * (_b[weight:_cons] + _b[weight:age]*31) * 31)
```

	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
Y1M1	116.7558	.1999727	583.86	0.000	116.3638	117.1477

```
. nlcom                                                                         ///
> (Y0M0:  _b[sbp:_cons]                                                         ///
>       + _b[sbp:age]*30                                                        ///
>       + _b[sbp:weight]         * (_b[weight:_cons] + _b[weight:age]*30)       ///
>       + _b[sbp:c.weight#c.age] * (_b[weight:_cons] + _b[weight:age]*30) * 30) ///
> (Y1M0:  _b[sbp:_cons]                                                         ///
>       + _b[sbp:age]*31                                                        ///
>       + _b[sbp:weight]         * (_b[weight:_cons] + _b[weight:age]*30)       ///
>       + _b[sbp:c.weight#c.age] * (_b[weight:_cons] + _b[weight:age]*30) * 31) ///
> (Y0M1:  _b[sbp:_cons]                                                         ///
>       + _b[sbp:age]*30                                                        ///
>       + _b[sbp:weight]         * (_b[weight:_cons] + _b[weight:age]*31)       ///
>       + _b[sbp:c.weight#c.age] * (_b[weight:_cons] + _b[weight:age]*31) * 30) ///
> (Y1M1:  _b[sbp:_cons]                                                         ///
>       + _b[sbp:age]*31                                                        ///
>       + _b[sbp:weight]         * (_b[weight:_cons] + _b[weight:age]*31)       ///
>       + _b[sbp:c.weight#c.age] * (_b[weight:_cons] + _b[weight:age]*31) * 31)
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
Y0M0	116.2711	.2027684	573.42	0.000	115.8737	116.6685
Y1M0	116.6976	.1999543	583.62	0.000	116.3057	117.0895
Y0M1	116.3306	.202873	573.42	0.000	115.933	116.7282
Y1M1	116.7558	.1999727	583.86	0.000	116.3638	117.1477

Different Treatment Effects

- Denoting E[Y(t,M(t'))] as $Y_{tM_{t'}}$, we define the following treatment effects of interest:

(Total) natural indirect effect (**NIE**)	$Y_{1M_1} - Y_{1M_0}$	δ(1)
(Pure) natural direct effect (**NDE**)	$Y_{1M_0} - Y_{0M_0}$	ζ (0)
(Pure) natural indirect effect (**PNIE**)	$Y_{0M_1} - Y_{0M_0}$	δ(0)
(Total) natural direct effect (**TNDE**)	$Y_{1M_1} - Y_{0M_1}$	ζ (1)
Total effect (**TE**)	$Y_{1M_1} - Y_{0M_0}$	τ

```
. mediate (sbp) (weight) (age, continuous(30 31)), all nolegend

Iteration 0:  EE criterion =  3.266e-25
Iteration 1:  EE criterion =  6.029e-28

Causal mediation analysis                              Number of obs = 7,299

Outcome model:     Linear
Mediator model:    Linear
Mediator variable: weight
Treatment type:    Continuous
```

sbp	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
Y0M0	116.2711	.1742669	667.20	0.000	115.9295	116.6126
Y1M0	116.6976	.1751862	666.13	0.000	116.3543	117.041
Y0M1	116.3306	.1740157	668.51	0.000	115.9895	116.6716
Y1M1	116.7558	.1747872	667.99	0.000	116.4132	117.0983
NIE						
age (1 vs 0)	.0581394	.0033109	17.56	0.000	.0516503	.0646286
NDE						
age (1 vs 0)	.4265509	.0092994	45.87	0.000	.4083243	.4447774
PNIE						
age (1 vs 0)	.0595155	.0032948	18.06	0.000	.0530577	.0659732
TNDE						
age (1 vs 0)	.4251748	.0092848	45.79	0.000	.4069769	.4433727
TE						
age (1 vs 0)	.4846903	.008329	58.19	0.000	.4683657	.5010148

Note: Outcome equation includes treatment-mediator interaction.

Total Natural Indirect Effect Y1M1 - Y1M0

Pure Natural Direct Effect Y1M0 - Y0M0

Pure Natural Indirect Effect Y0M1 - Y0M0

Total Natural Direct Effect Y1M1 - Y0M1

Total Effect Y1M1 - Y0M0

Which Decomposition?

- Practical question remains: For a specific analysis, which decomposition should be used? $\tau = \delta(1) + \zeta(0)$ or $\tau = \delta(0) + \zeta(1)$

- Or should both be used?

- We follow Nguyen et al. (2020) and propose three answers for three cases.

Which Decomposition?

Case 1: Is there a mediated effect? Or, is the causal effect partly mediated by this mediator?

- We propose using $\tau = \delta(1) + \zeta(0)$ decomposition (NIE and NDE)
- **Rational:** Here, we are not questioning the existence of a direct effect.
- We are researching the possibility of a mediated effect to the direct effect.
- If there is no mediated effect, then the total effect $\tau = \zeta(0)$ is the direct effect.

Which Decomposition?

```
. mediate (sbp) (weight) (age, continuous(30 31)), nie nde nolegend

Iteration 0:  EE criterion =  3.193e-25
Iteration 1:  EE criterion =  7.329e-29

Causal mediation analysis                              Number of obs = 7,299

Outcome model:     Linear
Mediator model:    Linear
Mediator variable: weight
Treatment type:    Continuous
--------------------------------------------------------------------------------
             |               Robust
         sbp | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+------------------------------------------------------------------
NIE          |
         age |
    (1 vs 0) |   .0581394   .0033109    17.56   0.000     .0516503    .0646286
-------------+------------------------------------------------------------------
NDE          |
         age |
    (1 vs 0) |   .4265509   .0092994    45.87   0.000     .4083243    .4447774
--------------------------------------------------------------------------------
Note: Outcome equation includes treatment-mediator interaction.
```

Which Decomposition?

Case 2: In addition to the mediated effect, is there a direct effect?

- We propose using $\tau = \delta(0) + \zeta(1)$ decomposition (PNIE and TNDE).
- This is a mirror image of the Case 1.
- **Rational:** Here, we are not questioning the existence of a mediator effect.
- We are researching the possibility of treatment affecting the outcome through other mechanisms.
- If there is no direct effect, then the total effect $\tau = \delta(0)$ is the indirect effect.

Which Decomposition?

```
. mediate (sbp) (weight) (age, continuous(30 31)), pnie tnde nolegend

Iteration 0:  EE criterion =  3.218e-25
Iteration 1:  EE criterion =  2.812e-28

Causal mediation analysis                          Number of obs = 7,299

Outcome model:     Linear
Mediator model:    Linear
Mediator variable: weight
Treatment type:    Continuous
```

sbp	Coefficient	Robust std. err.	z	P>|z|	[95% conf. interval]	
PNIE						
age						
(1 vs 0)	.0595155	.0032948	18.06	0.000	.0530577	.0659732
TNDE						
age						
(1 vs 0)	.4251748	.0092848	45.79	0.000	.4069769	.4433727

Note: Outcome equation includes treatment–mediator interaction.

Which Decomposition?

Case 3: No prior assumption or preferred question about either direct or indirect effect

- We propose reporting both $\tau = \delta(1) + \zeta(0)$ and $\tau = \delta(0) + \zeta(1)$ decompositions.

- **Rational:** If the purpose is to describe all we can learn, there is no reason to prefer wither decomposition over the other.

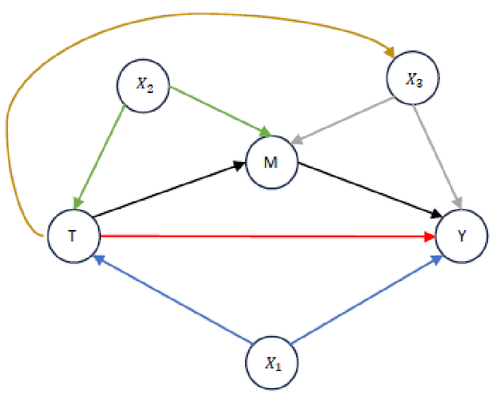
Which Decomposition?

```
. mediate (sbp) (weight) (age, continuous(30 31)), nie nde pnie tnde nolegend

Iteration 0:  EE criterion =  3.228e-25
Iteration 1:  EE criterion =  3.507e-28

Causal mediation analysis                                Number of obs = 7,299

Outcome model:     Linear
Mediator model:    Linear
Mediator variable: weight
Treatment type:    Continuous
```

sbp	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
age						
(1 vs 0)	.0581394	.0033109	17.56	0.000	.0516503	.0646286
NDE						
age						
(1 vs 0)	.4265509	.0092994	45.87	0.000	.4083243	.4447774
PNIE						
age						
(1 vs 0)	.0595155	.0032948	18.06	0.000	.0530577	.0659732
TNDE						
age						
(1 vs 0)	.4251748	.0092848	45.79	0.000	.4069769	.4433727

Note: Outcome equation includes treatment-mediator interaction.

Causal Identification: Assumptions



Sequential ignorabilty

1. No unobserved confounding in the treatment-outcome relationship.
2. No unobserved confounding in the mediator-outcome relationship.
3. No unmeasured confounding in the treatment-mediator relationship
4. No (observed) confounders in the mediator-outcome relationship that are caused by the treatment.

In addition to **sequential ignorability**, we need **SUTVA** and **overlap** assumptions.

Continuous Outcomes With `mediate`

```
. mediate (sbp i.healthstat calories)        ///
>         (weight i.healthstat calories)     ///
>         (age, continuous(30 31)), nolegend

Iteration 0:  EE criterion =  2.881e-26
Iteration 1:  EE criterion =  9.487e-29

Causal mediation analysis                                Number of obs = 4,876

Outcome model:      Linear
Mediator model:     Linear
Mediator variable:  weight
Treatment type:     Continuous
```

sbp	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
age						
(1 vs 0)	.0201403	.0026806	7.51	0.000	.0148864	.0253943
NDE						
age						
(1 vs 0)	.4048525	.0109781	36.88	0.000	.3833358	.4263692
TE						
age						
(1 vs 0)	.4249928	.0110645	38.41	0.000	.4033068	.4466788

Note: Outcome equation includes treatment-mediator interaction.

Continuous Outcomes With **`mediate`**

```
. estat proportion

Proportion mediated                                    Number of obs = 4,876

Continuous treatment levels:
  0: age = 30 (control)
  1: age = 31
```

sbp	Proportion	Robust std. err.	z	P>|z|	[95% conf. interval]	
age (1 vs 0)	.0473898	.0062387	7.60	0.000	.0351623	.0596174

Outline

- Causal Inference
  - ✓ – Treatment effects with **`regress`**
  - ✓ – Treatment effects with **`teffects ra`**
- Mediation
  - ✓ – Mediation using **`regress`**
  - ✓ – Mediation using **`sem`** and **`estat teffects`**
  - ✓ – Mediation using **`sem`** and **`bootstrap`**
- Causal Mediation
  - ✓ – Continuous outcomes with **`mediate`**
  - – Binary outcomes with **`mediate`**
  - – Count mediators with **`mediate`**

Binary Outcomes With **`mediate`**

```
. mediate (hypertension i.healthstat calories, logit)  ///
>         (hx_overweight i.healthstat calories, logit) ///
>         (age, continuous(30 31)), nolegend

Iteration 0:  EE criterion =  4.470e-21
Iteration 1:  EE criterion =  3.422e-32

Causal mediation analysis                              Number of obs = 4,459

Outcome model:     Logit
Mediator model:    Logit
Mediator variable: hx_overweight
Treatment type:    Continuous
```

hypertension	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
age (1 vs 0)	.0002823	.000064	4.41	0.000	.0001569	.0004077
NDE						
age (1 vs 0)	.0118771	.0003093	38.40	0.000	.0112708	.0124834
TE						
age (1 vs 0)	.0121594	.0002964	41.03	0.000	.0115785	.0127403

Note: Outcome equation includes treatment-mediator interaction.

Risk Ratios

- If the outcome is binary, and if the outcome model is either logit or probit, we can express the treatment effects as risk ratios or odds ratios.

- The treatment effects on risk-ratio are ratios of potential-outcome means:

$$NIE^{RR} = \frac{Y_{1M_1}}{Y_{1M_0}}$$
$$NDE^{RR} = \frac{Y_{1M_0}}{Y_{0M_0}}$$
$$PNIE^{RR} = \frac{Y_{0M_1}}{Y_{0M_0}}$$
$$TNDE^{RR} = \frac{Y_{1M_1}}{Y_{0M_1}}$$
$$TE^{RR} = \frac{Y_{1M_1}}{Y_{0M_0}}$$

Risk Ratios With `estat rr`

```
. estat rr
estat rr requires potential-outcome means; refitting model ...

Transformed treatment effects                        Number of obs = 4,459

Continuous treatment levels:
  0: age = 30 (control)
  1: age = 31
```

hypertension	Risk ratio	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
age (1 vs 0)	1.001121	.0002547	4.40	0.000	1.000622	1.00162
NDE						
age (1 vs 0)	1.049488	.002289	22.15	0.000	1.045011	1.053984
TE						
age (1 vs 0)	1.050664	.0022799	22.78	0.000	1.046205	1.055142

Odds Ratios

- For logit and probit outcome models, Y_{tM_t}, are probabilities, and so the treatment effects on the odds-ratio scale are

$$NIE^{OR} = \frac{Y_{1M_1}/(1 - Y_{1M_1})}{Y_{1M_0}/(1 - Y_{1M_0})}$$
$$NDE^{OR} = \frac{Y_{1M_0}(1 - Y_{1M_0})}{Y_{0M_0}(1 - Y_{0M_0})}$$
$$PNIE^{OR} = \frac{Y_{0M_1}(1 - Y_{0M_1})}{Y_{0M_0}(1 - Y_{0M_0})}$$
$$TNDE^{OR} = \frac{Y_{1M_1}(1 - Y_{1M_1})}{Y_{0M_1}(1 - Y_{0M_1})}$$
$$TE^{OR} = \frac{Y_{1M_1}(1 - Y_{1M_1})}{Y_{0M_0}(1 - Y_{0M_0})}$$

- Similar to risk-ratio, the total effect is the **product** of direct and indirect effect.

Odds Ratios With **estat or**

```
. estat or
estat or requires potential-outcome means; refitting model ...

Transformed treatment effects                          Number of obs = 4,459

Continuous treatment levels:
  0: age = 30 (control)
  1: age = 31
```

hypertension	Odds ratio	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
age (1 vs 0)	1.001499	.0003395	4.42	0.000	1.000833	1.002164
NDE						
age (1 vs 0)	1.066149	.0024518	27.85	0.000	1.061355	1.070966
TE						
age (1 vs 0)	1.067747	.0024108	29.03	0.000	1.063032	1.072483

Continuous Treatment With Multiple Values

```
. mediate (hypertension i.healthstat calories, logit)  ///
>         (hx_overweight i.healthstat calories, logit) ///
>         (age, continuous(30 31 35)), nolegend
```

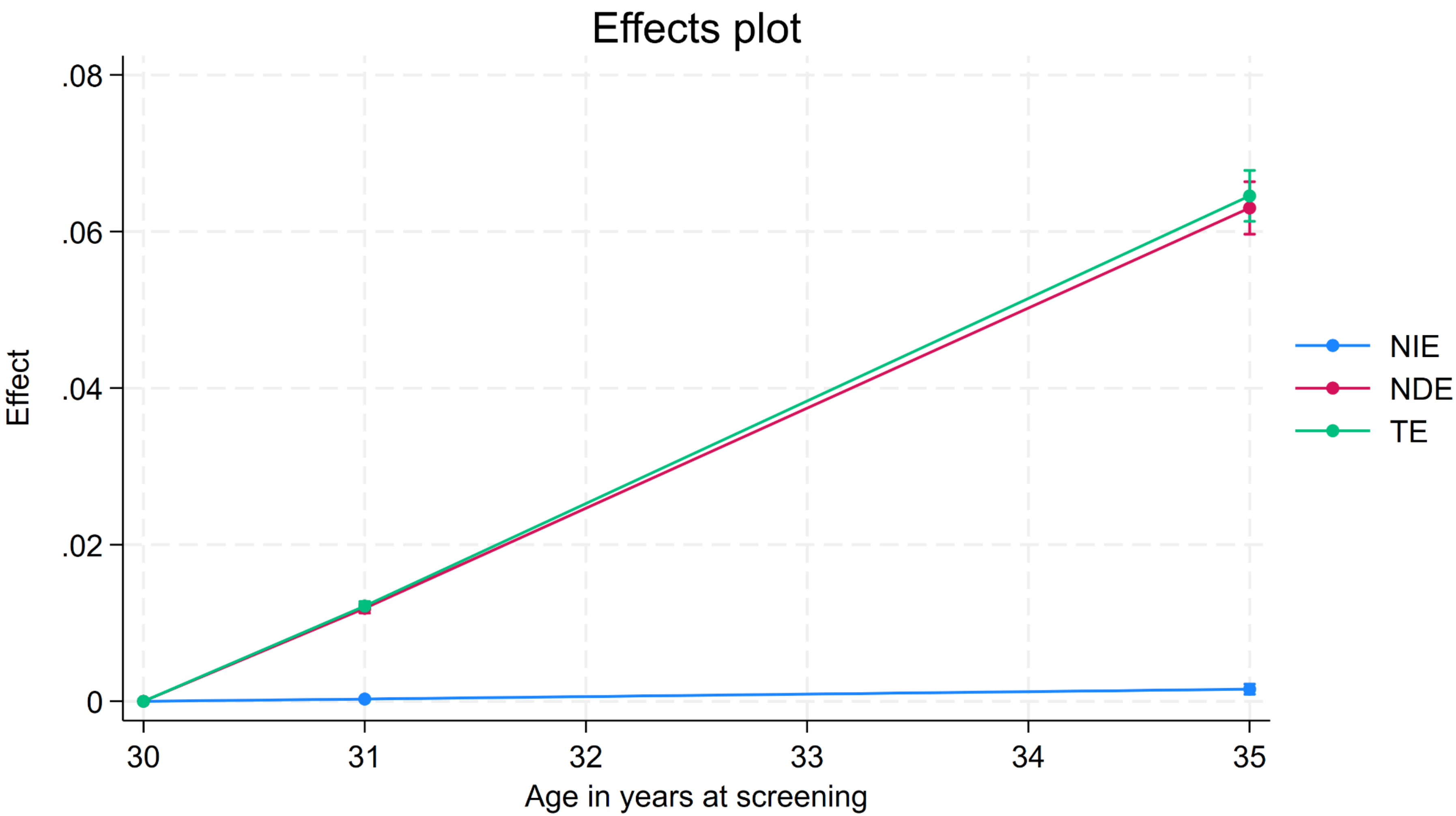
hypertension	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
age						
(1 vs 0)	.0002823	.000064	4.41	0.000	.0001569	.0004077
(2 vs 0)	.001543	.0003315	4.65	0.000	.0008933	.0021927
NDE						
age						
(1 vs 0)	.0118771	.0003093	38.40	0.000	.0112708	.0124834
(2 vs 0)	.0630108	.0017107	36.83	0.000	.0596579	.0663637
TE						
age						
(1 vs 0)	.0121594	.0002964	41.03	0.000	.0115785	.0127403
(2 vs 0)	.0645538	.0016592	38.91	0.000	.0613019	.0678057

Note: Outcome equation includes treatment-mediator interaction.

Continuous Treatment With Multiple Values

```
. estat or
estat or requires potential-outcome means; refitting model ...

Transformed treatment effects                          Number of obs = 4,459

Continuous treatment levels:
  0: age = 30 (control)
  1: age = 31
  2: age = 35

-------------------------------------------------------------------------------
             |               Robust
hypertension | Odds ratio   std. err.       z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
NIE          |
         age |
   (1 vs 0)  |   1.001499   .0003395     4.42   0.000     1.000833    1.002164
   (2 vs 0)  |   1.007322   .0015833     4.64   0.000     1.004224     1.01043
-------------+-----------------------------------------------------------------
NDE          |
         age |
   (1 vs 0)  |   1.066149   .0024518    27.85   0.000     1.061355    1.070966
   (2 vs 0)  |   1.376683   .0155514    28.30   0.000     1.346538    1.407503
-------------+-----------------------------------------------------------------
TE           |
         age |
   (1 vs 0)  |   1.067747   .0024108    29.03   0.000     1.063032    1.072483
   (2 vs 0)  |   1.386763   .0154809    29.29   0.000     1.356751     1.41744
-------------------------------------------------------------------------------
```

estat effectplot



Outline

- Causal Inference
  - ✓ – Treatment effects with **`regress`**
  - ✓ – Treatment effects with **`teffects ra`**
- Mediation
  - ✓ – Mediation using **`regress`**
  - ✓ – Mediation using **`sem`** and **`estat teffects`**
  - ✓ – Mediation using **`sem`** and **`bootstrap`**
- Causal Mediation
  - ✓ – Continuous outcomes with **`mediate`**
  - ✓ – Binary outcomes with **`mediate`**
  - – Count mediators with **`mediate`**

Count Mediators With **mediate**

```
. mediate (sbp) (smokenum, poisson) (age, continuous(30 31)), ///
>         nolog nolegend

Final EE criterion =  4.60e-29

Causal mediation analysis                              Number of obs = 1,035

Outcome model:     Linear
Mediator model:    Poisson
Mediator variable: smokenum
Treatment type:    Continuous
```

sbp	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
age (1 vs 0)	.0024086	.0066381	0.36	0.717	-.0106018	.015419
NDE						
age (1 vs 0)	.4576963	.0339078	13.50	0.000	.3912383	.5241544
TE						
age (1 vs 0)	.4601049	.0333714	13.79	0.000	.3946983	.5255116

Note: Outcome equation includes treatment–mediator interaction.

Count Mediators With **mediate**

```
. estat cde, mvalue(20)

Controlled direct effect                              Number of obs = 1,035

Mediator variable: smokenum
Mediator value = 20

Continuous treatment levels:
  0: age = 30 (control)
  1: age = 31
```

	CDE	Delta-method std. err.	z	P>|z|	[95% conf. interval]	
age (1 vs 0)	.4478632	.051657	8.67	0.000	.3466173	.549109

Outline

- Causal Inference
  - ✓ – Treatment effects with **`regress`**
  - ✓ – Treatment effects with **`teffects ra`**
- Mediation
  - ✓ – Mediation using **`regress`**
  - ✓ – Mediation using **`sem`** and **`estat teffects`**
  - ✓ – Mediation using **`sem`** and **`bootstrap`**
- Causal Mediation
  - ✓ – Continuous outcomes with **`mediate`**
  - ✓ – Binary outcomes with **`mediate`**
  - ✓ – Count mediators with **`mediate`**

References

Baron, RM & Kenny, DA. (1986) The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations . *Journal of Personality and Social Psychology* . Vol. 51, No. 6, 1173-1182

Hayes, AF. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis. Guilford: New York

MacKinnon DP, Kisbu-Sakarya Y, & Gottschall AC (2013). Developments in Mediation Analysis. In Little TD (editor), *The Oxford Handbook of Quantitative Methods*, Volume 2 (pp 361-386). Oxford: Oxford

References

Neal, B. (2020). Introduction to causal inference.

Nguyen, T., Schmid, I., and Stuart, E. (2020). Clarifying causal mediationanalysis for the applied researcher: Defining effects based on what we want to learn. Psychological Methods, 26.

Nguyen, Trang Quynh, Schmid, Ian, Ogburn, Elizabeth L. and Stuart, Elizabeth A.. "Clarifying causal mediation analysis: Effect identification via three assumptions and five potential outcomes" Journal of Causal Inference, vol. 10, no. 1, 2022, pp. 246-279.

VanderWeele, T. and Shpitser, I. (2013). On the definition of a confounder. The Annals of Statistics, 41.

Thank you!

Questions?

You can download the slides, datasets, and do-files here:
https://tinyurl.com/stata-mediation

You can contact me anytime at chuber@stata.com