Title

stata.com

mediate — Causal mediation analysis

Description

`mediate` fits causal mediation models and estimates effects of a treatment on an outcome. The treatment effect can occur both directly and indirectly through another variable, a mediator. The outcome and mediator variables may be continuous, binary, or count. The treatment may be binary, multivalued, or continuous. The estimated direct, indirect, and total effects have a causal interpretation provided that assumptions pertaining to causal mediation models are met.

Quick start

Fit the mediation model with continuous outcome `y1`, continuous mediator `m1`, and categorical treatment `t1`, and estimate the total effect, natural direct effect, and natural indirect effect

```
mediate (y1) (m1) (t1)
```

Same as above, but with covariates in both the outcome and the mediator equations

```
mediate (y1 x1 x2) (m1 x1 x3) (t1)
```

Same as above, but with probit model for binary outcome `y2` and Poisson model for count mediator `m2`

```
mediate (y2 x1 x2, probit) (m2 x1 x3, poisson) (t1)
```

Same as above, but estimate only the natural indirect effect (NIE)

```
mediate (y2 x1 x2, probit) (m2 x1 x3, poisson) (t1), nie
```

Same as above, but also estimate potential-outcome means

```
mediate (y2 x1 x2, probit) (m2 x1 x3, poisson) (t1), nie pomeans
```

Fit the mediation model with continuous treatment `t2`, and evaluate at values 0 and 4 of the treatment with 0 as the control

```
mediate (y2 x1 x2, probit) (m2 x1 x3, poisson) (t2, continuous(0 4))
```

Menu

Statistics > Causal inference/treatment effects > Continuous outcomes > Causal mediation

Statistics > Causal inference/treatment effects > Binary outcomes > Causal mediation

Statistics > Causal inference/treatment effects > Count outcomes > Causal mediation

Statistics > Causal inference/treatment effects > Nonnegative outcomes > Causal mediation

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.

omodel	Description
Model	
`linear`	linear model; the default
`expmean`	exponential-mean model
`logit`	logistic regression model
`probit`	probit regression model
`poisson`	Poisson model

omodel specifies the model for the outcome variable.

mmodel	Description
Model	
`linear`	linear model; the default
`expmean`	exponential-mean model
`logit`	logistic regression model
`probit`	probit regression model
`poisson`	Poisson model

mmodel specifies the model for the mediator variable.

The `logit` outcome model may not be combined with the `linear` or `expmean` mediator model; `probit` rather than `logit` may be used in these cases.

stat	Description
Stat	
Pearl's labeling of effects	
`nie`	natural indirect effect
`nde`	natural direct effect
`te`	total effect
`pnie`	pure natural indirect effect
`tnde`	total natural direct effect
ATE labeling of effects	
`aite`	average indirect treatment effect; synonym for `nie`
`adte`	average direct treatment effect; synonym for `nde`
`ate`	total 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`	all effects and potential-outcome means

Multiple effects may be specified; default is `nie nde te`.

options	Description
Model	
`nointeraction`	exclude interaction of mediator and treatment
`control(#` | *label*`)`	specify the level of *tvar* that is the control; default is first treatment level
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `robust`, `cluster` *clustvar*, `bootstrap`, or `jackknife`
`nose`	do not estimate standard errors
Reporting	
`level(#)`	set confidence level; default is `level(95)`
`ateterms`	use ATE terminology to label effects
`aequations`	display auxiliary-equation results
`nolegend`	suppress table legend
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Optimization	
optimization_options	control the optimization process; seldom used
Advanced	
`force`	force estimation when the number of treatment groups exceeds 10
`coeflegend`	display legend instead of statistics

omvarlist and *mmvarlist* may contain factor variables; see [U] **11.4.3 Factor variables**.

`bootstrap`, `by`, `collect`, `jackknife`, and `statsby` are allowed; see [U] **11.1.10 Prefix commands**.

Weights are not allowed with the `bootstrap` prefix; see [R] **bootstrap**.

`pweights`, `fweights`, and `iweights` are allowed; see [U] **11.1.6 weight**.

`coeflegend` does not appear in the dialog box.

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

Options

Model

`noconstant`; see [R] **Estimation options**.

`continuous(`*numlist*`)` specifies that the treatment variable is continuous; *numlist* specifies the values at which the potential-outcome means are to be evaluated, where the first value in the list is taken as the control.

`nointeraction` excludes the interaction between the treatment and the mediator; by default, the model includes the treatment–mediator interaction.

`control(`*#* | *label*`)` specifies the level of *tvar* that is the control. The default is the first treatment level. You may specify the numeric level *#* (a nonnegative integer) or the label associated with the numeric level. `control()` may not be specified with continuous treatments.

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`.

SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster *clustvar*), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [R] ***vce_option***.

nose suppresses calculation of the variance–covariance matrix and standard errors.

Reporting

level(#); see [R] **Estimation options**.

ateterms specifies that ATE terminology be used to label effects. ateterms is strictly a labeling option. This option may not be specified on replay.

aequations specifies that the estimation results for the outcome model and the mediator model be displayed. By default, they are not displayed.

nolegend suppresses the display of the table legend.

display_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(*style*), cformat(%*fmt*), pformat(%*fmt*), sformat(%*fmt*), and nolstretch; see [R] **Estimation options**.

Optimization

optimization_options: conv_maxiter(), conv_ptol(), conv_vtol(), tracelevel(), and [no]log. See [M-5] **optimize( )**.

conv_maxiter(#) specifies the maximum number of iterations. The default is the number set using set maxiter, which by default is 300.

conv_ptol(#) specifies the convergence criteria for the parameters. The default is conv_ptol(1e-6).

conv_vtol(#) specifies the convergence criteria for the gradient. The default is conv_vtol(1e-7).

tracelevel(*tracelevel*) allows you to display additional information about the iterative process in the iteration log. *tracelevel* may be none, value, tolerance, step, params, or gradient. See *tracelevel* in [M-5] **optimize( )** for details.

log and nolog specify whether to display the iteration log. The iteration log is displayed by default unless you used set iterlog off to suppress it; see set iterlog in [R] ***set iter***.

Advanced

force forces estimation when the number of treatment groups exceeds 10. By default, only 10 groups are allowed for multivalued treatments. Do not use the force option if the treatment is continuous; instead, use the continuous() option.

The following option is available with mediate but is not shown in the dialog box:

coeflegend; see [R] **Estimation options**.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Introduction

Causal inference is an essential goal in many research areas and aims at identifying and quantifying causal effects. For example, we might wish to find out whether physical exercise leads to an improvement in self-perceived well-being, and if so, to what extent. Causality in this context typically means that there is some cause T that has an effect on some outcome Y. We could visualize this relation with a simple causal diagram:

Figure 1

If T is a measure of exercise and Y is well-being, then under certain assumptions, we could use the above causal model to identify the total effect of exercise on well-being (by means of a randomized controlled trial, for instance). However, a question that we cannot answer empirically with our simple causal model is why exercise may increase well-being. Perhaps exercising causes an increase in certain chemicals or hormones in the human body, which in turn affects perceptions of well-being. To assess such intermediary effects, we need to expand our simple causal model by adding variables that lie on the causal pathway between T and Y:

Figure 2

Suppose that, in our exercise example, the variable M represents the production of a certain chemical in the human body. With this new model, we now hypothesize that exercising leads to the production of this chemical, which in turn leads to an increase in well-being. However, it might be unrealistic to assume that the effect of exercise on well-being hinges exclusively on the production

of that chemical. Perhaps we would like to allow for the possibility that exercise has an effect on well-being beyond its path through the mediating variable, and so a better model might be

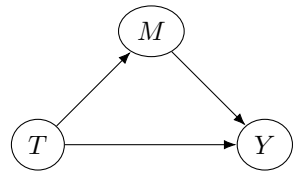


Figure 3

Here, we include a direct path from T to Y in addition to the indirect path of T to Y via M. In other words, we assume that exercise produces a particular chemical that affects well-being, but we also allow for the possibility of a direct effect of exercise on well-being that is not related to the chemical. This is the classical mediation model that decomposes the total effect into a direct and an indirect effect. Causal mediation analysis aims to identify these direct and indirect effects and give them a causal interpretation.

Approaches to mediation analysis

Mediation analysis can be performed in a variety of ways. The classical approach of Baron and Kenny (1986) fits two linear regression models, one for M and one for Y, and estimates direct, indirect, and total effects as functions of the coefficients. Estimation can be simplified by fitting the models for M and Y simultaneously via structural equation modeling as discussed in [SEM] **Example 42g**. In Stata, you can use `sem` to fit linear models for the outcome and mediator, and you can then use `estat teffects` to obtain a decomposition of direct and indirect effects based on the results from `sem`. Note that this classical approach relies on the specification of a particular model at the outset of the process.

Another approach to mediation analysis is based on the potential-outcomes framework. The potential outcomes are values of the outcome that would be obtained under different conditions, such as when the treatment occurs. Differences in potential outcomes yield direct, indirect, and total effects of interest. This is the approach typically referred to as causal mediation analysis and is the one implemented in `mediate`.

The causal mediation framework allows much flexibility. In this framework, it is common to allow the mediator and the treatment to interact; thus, we do not assume that the effect of the mediator on the outcome is the same for the treated and untreated groups. The total effect of the treatment on the outcome can be decomposed into direct and indirect effects in two ways, and the researcher can select the decomposition that matches his or her research question. The effects are defined in a model-free manner, so the researcher can select an estimation method that is appropriate for his or her data and then compute estimates of the effects of interest.

In the situation where both the outcome and the mediator are modeled using linear regression and there is no treatment–mediator interaction, the classical approach and causal mediation via the potential-outcomes framework will lead to the same results.

Workflow for causal mediation

The general workflow for researchers performing causal mediation analysis is as follows:

1. Specify your research question.
2. Identify the treatment, mediator, and outcome variables to be analyzed.
3. Determine which effect decomposition can be used to answer your research question.

4. Evaluate whether assumptions for causal interpretation are appropriate.
5. Select a method for estimating the causal effects of interest.
6. Interpret the results.

In our introductory discussion, we provided an example of step 2 by using exercise, chemical production, and well-being as the treatment, mediator, and outcome variables, respectively. Below, we will provide a conceptual introduction to steps 1, 3, 4, and 5. In *Examples*, we will interpret the results in different scenarios.

Forming research questions

Before performing causal mediation analysis, we must decide which research questions motivated the desire to perform the analysis. Here are some types of research questions that may arise:

1. A scenario in which the primary interest is to determine whether there is an indirect effect and, if so, to quantify it. In our example above, we might assume there will be some direct impact of exercise on well-being, but we also wonder if and to what extent there is an indirect effect, such as exercise increasing production of a chemical which in turn increases well-being. In this case, we would be interested in decomposing the effects according to ATE decomposition 1.
2. A scenario in which the primary interest is to determine whether there is a direct effect and, if so, to quantify it. Continuing with our example, perhaps we expect an indirect effect, but we also wish to determine if there are any other ways in which exercise causes changes in well-being. In this case, we would be interested in decomposing the effects according to ATE decomposition 2.
3. A scenario in which the primary interest is to determine how the total effect can be decomposed into direct and indirect effects, with focus remaining on all effects and not just direct or just indirect effects. In our example, we simply want to explore the breakdown of the total effect of exercise on well-being into all possible direct and indirect effects. In this case, we would likely be interested in looking at both decompositions, ATE decomposition 1 and ATE decomposition 2.
4. A scenario in which we want to determine the effect of the treatment on the outcome when the mediator is set to a specific value. In our example, we might want to know the effect of exercise on well-being for individuals whose level of this particular chemical is 10, which is the mean value in the population. In this case, we would be interested in controlled direct effects.

Potential outcomes and effect decompositions

Below, we introduce statistics that may be of interest when performing causal mediation analysis. Many of these statistics have a variety of names in the causal mediation literature. See, for instance, Robins and Greenland (1992), VanderWeele (2015), and Pearl and MacKenzie (2018) for some of the various terminology.

1. **Potential-outcome means.** These estimate the population-average value of the outcome that would be expected if everyone was given the treatment (denoted here as $Y[1, M(1)]$) or if everyone was given the control (denoted $Y[0, M(0)]$). In our example, $Y[1, M(1)]$ is the expected average well-being if everyone exercises, and $Y[0, M(0)]$ is the expected average well-being if no one exercises.

   In addition, there are two cross-world potential outcomes. These are a bit less intuitive because they correspond to situations that do not exist for any individual in the population. The first is

the expected value of the outcome when everyone is treated but counterfactually experiences the value of the mediator associated with being untreated (denoted $Y[1, M(0)]$). The second is the expected value of the outcome when everyone is untreated but counterfactually experiences the value of the mediator associated with being treated (denoted $Y[0, M(1)]$). In our example, $Y[1, M(0)]$ is the expected well-being if everyone was treated but experiencing the chemical level as if untreated. $Y[0, M(1)]$ is the expected value if everyone was untreated but experiencing the chemical level as if treated.

The `mediate` command reports these potential-outcome means when the `pomeans` option is specified.

2. **Total effect (TE).** This estimates the average difference in outcomes that we expect when everyone receives the treatment versus when no one receives the treatment. In our case, it estimates the improvement in well-being that we would expect if everyone exercises versus if no one exercises.

   The total effect is also referred to as the average treatment effect (ATE), the total average treatment effect, or the marginal total effect.

   The `mediate` command reports this statistic when the `te` option or its synonym `ate` is specified.

   The total effect can be decomposed into direct and indirect effects in two ways when we allow for a treatment–mediator interaction.

Decomposition 1. This decomposition separates the direct effect under the untreated mediator condition from the total indirect effect. Nguyen, Schmid, and Stuart (2021) recommend using this decomposition when a direct effect is assumed and the researcher is questioning whether a mediation effect also exists. In our example, we would be interested in this decomposition if we expect that exercise has a direct effect on well-being but want to determine whether a portion of the total effect can be attributed to the increase in the chemical (and if so, how much of the total effect is due to this mediation effect).

3. **Natural direct effect (NDE).** This estimates the average direct effect of the treatment on the outcome when the mediator is held at its value associated with being untreated. It is the difference $Y[1, M(0)] - Y[0, M(0)]$.

   This effect is sometimes referred to as the pure natural direct effect or the average direct treatment effect (ADTE). All remaining effects of the treatment on the outcome are included in the natural indirect effect.

   The `mediate` command reports this statistic when the `nde` option or its synonym `adte` is specified.

4. **Natural indirect effect (NIE).** This estimates the average indirect effect through a mediator. It is the difference $Y[1, M(1)] - Y[1, M(0)]$.

   This effect is sometimes referred to as the total natural indirect effect, causal mediation effect, or average indirect treatment effect (AITE).

   The `mediate` command reports this statistic when the `nie` option or its synonym `aite` is specified.

Decomposition 2. This decomposition separates the indirect effect under the untreated condition from the total direct effect. Nguyen, Schmid, and Stuart (2021) recommend using this decomposition when an indirect effect is assumed and the researcher is questioning whether a direct effect also exists. In our example, we would be interested in this decomposition if we believe that exercise increases production of the chemical which in turn increases well-being but want to determine if there is also some change in well-being that is not caused by this mediation effect (and if so, how much of the total effect is not due to the mediation effect).

5. **Pure natural indirect effect (PNIE).** This estimates the average indirect effect of a mediator under the untreated/control condition. It is the difference $Y[0, M(1)] - Y[0, M(0)]$.

   This is sometimes referred to as the average indirect treatment effect with respect to controls (AITEC). All remaining effects of the treatment on the outcome are included in the total natural direct effect.

   The `mediate` command reports this statistic when the `pnie` option or its synonym `aitec` is specified.

6. **Total natural direct effect (TNDE).** This estimates the average direct treatment effect when the mediator is held at its value associated with being treated. It is the difference $Y[1, M(1)] - Y[0, M(1)]$.

   This effect is sometimes referred to as the average direct treatment effect with respect to the treated (ADTET).

   The `mediate` command reports this statistic when the `tnde` option or its synonym `adtet` is specified.

When no prior assumptions are made about the existence of direct and indirect effects, Nguyen, Schmid, and Stuart (2021) recommend reporting both Decomposition 1 and Decomposition 2.

7. **Controlled direct effects.** These are the direct effects when the mediator is controlled by setting it to a specific value. After fitting your model with `mediate`, you can estimate the average controlled direct effect with the mediator set to your selected value by using `estat cde`; see *Example 11: Estimating controlled direct effects*. In our well-being example, controlled direct effects provide the direct effect of exercise on well-being when the chemical is assumed to be a specific value.

Evaluating assumptions for causal inference

Before proceeding to estimation and interpretation of the effects of interest, we need to verify that it is reasonable to give them a causal interpretation in our particular research context.

General assumptions for causal inference are discussed in [CAUSAL] **Intro**, and more precise definitions in the context of mediation are provided in *Assumptions for causal identification* below.

To evaluate whether assumptions of causality are met for our mediation model, we must first consider all potential variables, both observed and unobserved, that could affect the relationships among our treatment, mediator, and outcome. If we anticipate that there are confounders (variables that affect both an outcome and a predictor), we must determine whether these confounders will lead to biased results in the estimation of effects from our mediation analysis. In particular, we want to assume that

1. There is no unobserved confounding in the treatment–outcome relationship, and observed confounders are included as covariates in the outcome model.
2. There is no unobserved confounding in the mediator–outcome relationship, and observed confounders are included as covariates in the outcome model.
3. There is no unmeasured confounding in the treatment–mediator relationship, and observed confounders are included as covariates in the mediator model.
4. There are no confounders in the mediator–outcome relationship that are caused by the treatment. No variable exists that affects both the mediator and the outcome and that itself is caused by the treatment.

Estimation of effects

When assumptions are met, the `mediate` command can be used to estimate the causal parameters of interest.

While the effects derived under the potential-outcomes framework required no particular model, we now need to decide how to model our data to obtain estimates.

We first select models for the outcome and the mediator. Outcomes can be continuous, binary, or counts and can be modeled using a linear, exponential-mean, logistic, probit, or Poisson model. Mediators can also be continuous, binary, or counts and can also be modeled using a linear, exponential-mean, logistic, probit, or Poisson model. Covariates can be included in the outcome and mediator models. The treatment may be binary, categorical (multivalued), or continuous.

As a simple example of the `mediate` command, say that we have a binary outcome `y`, a continuous mediator `m`, and a binary treatment `t`. We can fit a mediation model by typing

```
. mediate (y, probit) (m, linear) (t)
```

The first set of parentheses specifies a model for the outcome. The second set of parentheses specifies the model for the mediator. The third set of parentheses defines the treatment. By default, the TE and its decomposition into NDE and NIE are reported in the output.

If we would instead like to see the second type of decomposition, we can obtain the TE, PNIE, and TNDE by typing

```
. mediate (y, probit) (m, linear) (t), te pnie tnde
```

Many combinations of models and effects can be obtained. See *Examples* below for additional syntax examples as well as interpretation of the results.

Technical overview of causal mediation

Above, we provided a conceptual introduction to the concepts in causal mediation analysis. Here we more formally define the potential-outcomes framework; explain total, direct, and indirect effects; and introduce the assumptions necessary for causal inference. If you are familiar with the technical aspects of causal mediation and are ready to see `mediate` demonstrated, go directly to *Examples*.

Mediation analysis in the potential-outcomes framework

The potential-outcomes framework is commonly employed for identifying causal effects. If we go back to the model associated with figure 1 and assume T is a binary treatment, we can identify two sets of potential outcomes, $Y_i(1)$ and $Y_i(0)$. $Y_i(t)$ is the outcome that would be realized if the ith individual were exposed to treatment level t.

Consider a randomized experiment where the experimental group exercises while the control group spends the same amount of time in a resting state. The outcome is subjective well-being that is measured after exercising/resting. If it were possible to observe an individual in both states at the same time, we would observe one outcome value under treatment, $Y_i(1)$, and one value under the control condition, $Y_i(0)$. Then the treatment effect would be the difference $\tau_i = Y_i(1) - Y_i(0)$. In other words, there is a potential outcome for each treatment level that could be administered. Averaging the difference over all individuals in the sample would yield an estimate of the ATE $\tau = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)]$.

However, it is not possible to observe the same individual under both conditions at the same time; we can only observe one of these while the other is missing. If an individual is treated, we observe $Y_i(1)$, and if not, we observe $Y_i(0)$. This has been coined the "fundamental problem of causal inference" (Holland 1986). Much of the treatment effects and causal inference literature deals with the question of how to estimate an ATE in the presence of this problem.

In a simple experiment where treatment is randomly assigned, the potential outcomes are independent of treatment assignment and the missing potential outcomes are missing completely at random. In this case, the average of the treatment group outcomes are a valid estimate of $E[Y_i(1)]$, and the average of the control group outcomes are a valid estimate of $E[Y_i(0)]$. Then $\widehat{\tau} = \widehat{E}[Y_i(1)] - \widehat{E}[Y_i(0)]$ where $\widehat{E}[Y_i(t)] = 1/N_t \sum_{i=1}^{N_t} 1(T_i = t)Y_i$ is a valid estimator of the ATE. This estimation strategy follows from the identification result that $E[Y_i(t)] = E(Y_i|T_i = t)$ such that $\tau = E[Y_i(1)] - E[Y_i(0)] = E[Y_i|T_i = 1] - E[Y_i|T_i = 0]$.

With observational rather than experimental data, however, the potential outcomes are not independent of the treatment assignment process, and the causal effect is not identifiable without imposing further assumptions such as conditional independence. Stata's `teffects` suite of commands provides a variety of estimators from this class of treatment-effects estimators.

For further information about identification and estimation in the context of causal models as well as an overview of estimators implemented in Stata, see [CAUSAL] **Intro**. Here we focus on causal inference and potential outcomes specifically for mediation analysis. In this situation, we have another set of potential outcomes, $M_i(1)$ and $M_i(0)$, because M is also affected by the treatment. That is, we can only observe $M_i(1)$ for the group of individuals who were treated, and we can only observe $M_i(0)$ for the controls. If we let t denote the treatment level with respect to the outcome and let t' be the treatment level with respect to the mediator, then the potential outcomes become $Y_i[t, M_i(t')]$.

Similar to the nonmediation case above, we can define a treatment effect as a difference between potential outcomes. The treatment effect is identified if

$$E[Y_i(t, M_i(t'))] = E_{M_i|T_i=t'} E[Y_i|M_i, T_i = t]$$

where $E_{M_i|T_i=t'}$ is the expectation of the mediator conditional on the treatment taking on the value t' and where $E[Y_i|M_i, T_i = t]$ is the expectation of the outcome conditional on the mediator and treatment taking on the value t.

Total, direct, and indirect effects

In mediation analysis, we are interested not only in the total treatment effect but also in its decomposition into direct effects and indirect effects.

Notice that if $t = t'$ for a given potential outcome, the resulting potential outcome is equivalent to $Y_i(t)$. Assuming again a binary treatment, we have that

$$\tau = E[Y_i(1)] - E[Y_i(0)] = E[Y_i(1, M_i(1))] - E[Y_i(0, M_i(0))]$$

In the context of mediation analysis, this treatment effect is also referred to as the total effect.

The total effect can be decomposed further into direct and indirect effects using contrasts between potential-outcome means. The contrasts yielding direct and indirect effects use potential outcomes for which $t \neq t'$, which means we set the treatment level to t and set the mediator to its potential value under treatment level t'.

The natural indirect effect is then defined as

$$\delta(t) \equiv E[Y_i(t, M_i(1))] - E[Y_i(t, M_i(0))], \quad t \in \{0, 1\}$$

Notice that here we "switch" the treatment from on to off in its effect on the mediator but keep the treatment fixed at value t in its effect on the outcome. This natural indirect effect is also sometimes referred to as the causal mediation effect (Imai, Keele, and Tingley 2010).

Likewise, the natural direct effect can be defined as

$$\zeta(t) \equiv E[Y_i(1, M_i(t))] - E[Y_i(0, M_i(t))], \quad t \in \{0, 1\}$$

Comparison of potential outcomes and classical mediation analysis

For those familiar with classical mediation analysis for linear models, it may be helpful to see how the calculation of total, direct, and indirect effects in the potential-outcomes framework relates to the classical product-of-coefficients approach.

We first write our mediation model corresponding with figure 3 as

$$Y_i = \beta_0 + \beta_1 M_i + \beta_2 T_i + \epsilon_i$$

$$M_i = \alpha_0 + \alpha_1 T_i + \nu_i$$

where ϵ_i and ν_i are uncorrelated error terms with means 0 and variances σ^2_ϵ and σ^2_ν, respectively.

Let's consider the indirect effect $\delta(1)$. To calculate $\delta(1)$ in the potential-outcomes framework, we need estimates for the potential-outcome means $E[Y_i(1, M_i(1))]$ and $E[Y_i(1, M_i(0))]$. Intuitively, what we want is a world where everyone in the population is exposed to the treatment, that is, $Y_i(1)$, but where we can switch the treatment on and off in regard to the effect of the treatment on the mediator, that is, $M_i(1)$ and $M_i(0)$. The difference when going from the treatment switched on to the treatment switched off will inform us about the effect of the treatment on the outcome that goes through the mediator. First, we write the above model in reduced form:

$$\begin{aligned} E[Y_i|M_i, T_i] &= \beta_0 + \beta_1(\alpha_0 + \alpha_1 T_i) + \beta_2 T_i \\ &= \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1 T_i + \beta_2 T_i \end{aligned}$$

This yields the conditional expectation $E[Y_i|M_i, T_i]$ that we can observe from the data.

To obtain the potential-outcome means, we can modify the reduced-form model by replacing M_i with the expectation of M_i that we would observe if T_i had taken on the value t' for every unit in the population. That is,

$$E[Y_i(t, M_i(t'))] = \beta_0 + \beta_1 E[M_i(t')] + \beta_2 t, \quad t \in \{0, 1\}$$

Thus, to compute the potential-outcome mean $E[Y_i(1, M_i(1))]$, we must set the treatment T_i to 1 in both the outcome and the mediator equations. In other words, we fix both t and t' at 1:

$$\begin{aligned} E[Y_i(1, M_i(1))] &= \beta_0 + \beta_1 E[M_i(t')] + \beta_2 t, \quad t = t' = 1 \\ &= \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1 \times 1 + \beta_2 \times 1 \\ &= \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1 + \beta_2 \end{aligned}$$

However, to compute $E[Y_i(1, M_i(0))]$, we need to set treatment T_i to 1 in the outcome equation and need to set it to 0 in the mediator equation. Specifically, we fix $t' = 0$ and $t = 1$:

$$\begin{aligned} E[Y_i(1, M_i(0))] &= \beta_0 + \beta_1 E[M_i(t')] + \beta_2 t, \quad t = 1;\ t' = 0 \\ &= \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1 \times 0 + \beta_2 \times 1 \\ &= \beta_0 + \beta_1\alpha_0 + \beta_2 \end{aligned}$$

Calculating the difference between these two potential-outcome means yields the indirect treatment effect

$$\begin{aligned} \delta(1) &= (\beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1 + \beta_2) - (\beta_0 + \beta_1\alpha_0 + \beta_2) \\ &= \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1 + \beta_2 - \beta_0 - \beta_1\alpha_0 - \beta_2 \\ &= \beta_1\alpha_1 \end{aligned}$$

In this case of a linear model, the indirect treatment effect is the product of the treatment coefficient from the mediator equation and the mediator coefficient from the outcome equation. This is congruent with the indirect effect definition in the product-of-coefficients method for mediation as proposed by the classical mediation literature; see Baron and Kenny (1986).

Accounting for treatment–mediator interaction

Notice that the indirect effect we estimated above would be the same if we had estimated $\delta(0)$ instead. Thus far, we assumed that the effect of the mediator on the outcome is the same for both treatment groups. Presumably, a more realistic assumption would be to allow the mediator effects to vary by treatment. This can be achieved by including a treatment–mediator interaction term.

When we allow an interaction, $\delta(0) \neq \delta(1)$. Now we have two indirect effects, one with respect to treatment $[\delta(1)]$ and one with respect to controls $[\delta(0)]$. In the following, we will refer to $\delta(1)$ as the NIE and to $\delta(0)$ as the PNIE.

To illustrate computation of the NIE under inclusion of a treatment–mediator interaction, we write a new model

$$Y_i = \beta_0 + \beta_1 M_i + \beta_2 T_i + \beta_3 M_i T_i + \epsilon_i$$

$$M_i = \alpha_0 + \alpha_1 T_i + \nu_i$$

Here, NIE $\equiv E[Y_i(1, M_i(1)) - Y_i(1, M_i(0))]$, whereas PNIE $\equiv E[Y_i(0, M_i(1)) - Y_i(0, M_i(0))]$.

As before, to calculate NIE, we need potential-outcome means $E[Y_i(1, M_i(1))]$ and $E[Y_i(1, M_i(0))]$. Writing the model in reduced form, we get

$$\begin{aligned} E[Y_i|M_i, T_i] &= \beta_0 + \beta_2 T_i + \beta_1(\alpha_0 + \alpha_1 T_i) + \beta_3 T_i(\alpha_0 + \alpha_1 T_i) \\ &= \beta_0 + \beta_2 T_i + (\beta_1 + \beta_3 T_i)(\alpha_0 + \alpha_1 T_i) \end{aligned}$$

Fixing the values for the treatment in both equations accordingly, we have potential-outcome means

$$\begin{aligned} E[Y_i(1, M_i(1))] &= \beta_0 + \beta_2 \times 1 + (\beta_1 + \beta_3 \times 1)(\alpha_0 + \alpha_1 \times 1) \\ &= \beta_0 + \beta_2 + (\beta_1 + \beta_3)(\alpha_0 + \alpha_1) \end{aligned}$$

and

$$\begin{aligned} E[Y_i(1, M_i(0))] &= \beta_0 + \beta_2 \times 1 + (\beta_1 + \beta_3 \times 1)(\alpha_0 + \alpha_1 \times 0) \\ &= \beta_0 + \beta_2 + (\beta_1 + \beta_3)\alpha_0 \end{aligned}$$

Taking the difference yields the NIE

$$E[Y_i(1, M_i(1))] - E[Y_i(1, M_i(0))] = (\beta_1 + \beta_3)\alpha_1$$

We could proceed similarly for the other direct and indirect treatment effects. In this case with treatment–mediator interaction, we also have two direct treatment effects. We have NDE $\equiv E[Y_i(1, M_i(0)) - Y_i[0, M_i(0))]$ and TNDE $\equiv E[Y_i(1, M_i(1)) - Y_i(0, M_i(1))]$.

Notice that both treatment-effect decompositions—NIE and NDE as well as PNIE and TNDE—sum to the total treatment effect (or, as we will call it, the TE).

$$\text{TE} \equiv E[Y_i(1, M_i(1)) - Y_i(0, M_i(0))]$$

For further details and discussion on the different direct and indirect effects, as well as a discussion on the differences between causal inference and traditional mediation approaches, see Nguyen, Schmid, and Stuart (2021).

Assumptions for causal identification

The above discussion shows that the estimands of interest are the result of contrasts between potential-outcome means, which are conditional expectations of the outcome with respect to counterfactuals in both the outcome and the mediator equations. In other words, once we can estimate $E[Y_i(t, M_i(t'))]$, we can estimate all direct and indirect treatment effects of interest.

The general form for causal mediation potential-outcome means, which includes covariates X_i, can be written as the integral of the conditional expectation of the outcome with respect to the conditional distribution of the mediator (see Imai, Keele, and Tingley [2010]):

$$E[Y_i(t, M_i(t'))|X_i = x] = \int E[Y_i|M_i = m, T_i = t, X_i = x]\, dF[m|T_i = t', X_i = x]$$

This is a general, nonparametric solution that applies regardless of the underlying outcome and mediator models. Stata's `mediate` command uses analytical solutions of this integral for a variety of parametric outcome and mediator model combinations. Also, while so far we assumed a binary treatment for simplicity purposes, this approach generalizes straightforwardly to multivalued as well as continuous treatments.

As is the case with nonmediation causal inference, there are assumptions to be met for the estimated effects to be given a causal interpretation. Most notably, a crucial assumption in the nonmediation case is the conditional independence assumption, also known as conditional ignorability assumption, unconfoundedness, or selection on observables. This assumption states that potential outcomes are independent of treatment assignment after conditioning on a set of observed covariates that affect both the outcome and the selection into treatment (see Imbens [2004]). Intuitively, we have a model that resembles an experiment once we account for observable characteristics. More formally, we have that

$$Y_i(t) \perp T_i|X_i$$

In the mediation case, however, we have an additional selection process because "selection" into the mediator is also typically not based on random assignment. This leads to the following two conditional independence assumptions:

$$\{Y_i[t,m], M_i(t')\} \perp T_i | X_i = x$$
$$Y_i[t,m] \perp M_i(t') | T_i = t', X_i = x$$

The first assumption states that treatment assignment is independent of potential outcomes and potential mediators after conditioning on observed (pretreatment) covariates, or confounders. The second assumption states that potential mediators are independent of the potential outcomes given the observed treatment and observed (pretreatment) covariates. Because these assumptions are being made sequentially, this has also been coined the sequential ignorability assumption (Imai, Keele, and Tingley 2010).

Similarly, there is an additional overlap assumption with causal mediation models. In the nonmediation case, the overlap assumption states that each individual has a positive probability of receiving each treatment:

$$0 < \Pr(T_i = t | X_i = x), \quad t \in \{0, 1\}$$

In the mediation case, the same principle applies to the mediator:

$$0 < p(M_i(t) = m | T_i = t, X_i = x), \quad t \in \{0, 1\}$$

Finally, as is the case with nonmediation treatment-effects models, causal mediation models rely on the stable unit treatment-value assumption, which states that potential outcomes do not depend on treatments assigned to other individuals. For a detailed overview of effect identification and assumptions for causal mediation analysis, see Nguyen et al. (2022).

Examples

Example 1: A simple causal mediation model

Suppose we wish to find out whether exercise affects perceptions of well-being among some population of individuals. To the extent that there is such a causal relationship, we also wish to find out why exercise affects well-being.

We have fictional data from a randomized controlled trial with individuals randomized into two groups—one group performs physical exercise and the other group spends the same amount of time in a resting state. Subjective well-being is measured before and after treatment sessions. In addition, the level of the (fictional) hormone bonotonin is measured. The researchers wish to determine whether exercise leads to an increase in bonotonin levels, which in turn has a positive effect on subjective well-being. Here is an excerpt from our dataset:

```
. use https://www.stata-press.com/data/r18/wellbeing
(Fictional well-being data)
. list wellbeing bonotonin exercise age gender in 1/5, abbreviate(12)
```

	wellbeing	bonotonin	exercise	age	gender
1.	71.73816	196.5467	Control	58	Male
2.	68.66573	195.8572	Exercise	38	Female
3.	71.05155	228.6035	Exercise	53	Female
4.	69.44469	206.6651	Exercise	44	Female
5.	75.62035	261.6855	Exercise	28	Female

To estimate the treatment effects with `mediate`, we specify `wellbeing` as the outcome variable in the first set of parentheses, `bonotonin` as the mediator variable in the second set of parentheses, and `exercise` as the binary treatment variable in the third set of parentheses. Although inclusion of a treatment–mediator interaction is commonly recommended, we specify the `nointeraction` option here to omit the interaction and fit the simplest model possible.

```
. mediate (wellbeing) (bonotonin) (exercise), nointeraction
Iteration 0:  EE criterion =  1.627e-25
Iteration 1:  EE criterion =  3.061e-28
Causal mediation analysis                          Number of obs = 2,000
Outcome model:     Linear
Mediator model:    Linear
Mediator variable: bonotonin
Treatment type:    Binary
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
exercise (Exercise vs Control)	9.694617	.377312	25.69	0.000	8.955099	10.43413
NDE						
exercise (Exercise vs Control)	2.996658	.2109357	14.21	0.000	2.583231	3.410084
TE						
exercise (Exercise vs Control)	12.69127	.4005769	31.68	0.000	11.90616	13.47639

```
Note: Outcome equation does not include treatment-mediator interaction.
```

In the header of the output, we see that `mediate` fit linear models (the default) for both the outcome and the mediator. Three treatment-effect estimates are reported in the table. TE is the total effect of exercise on well-being and is estimated to be 12.7. The interpretation is the same as for the ATE in the nonmediation case: if everyone in the population exercised, their well-being would be, on average, 12.7 points higher than their well-being would be if no one exercised. The decomposition of the TE into direct and indirect effects is of primary interest. The NIE is estimated to be 9.7, whereas the NDE is estimated to be 3.0. These sum to the total effect of 12.7. The indirect effect is much larger than the direct effect, indicating that the effect of exercise on well-being is largely due to exercise affecting bonotonin levels, which in turn affect well-being. The direct effect of 3.0 is the effect of exercise on well-being beyond the effect through bonotonin.

Instead of comparing estimates of the direct and indirect effects, we might ask what proportion of the total effect is due to mediation. We can answer this question by using `estat proportion`.

```
. estat proportion
Proportion mediated                                     Number of obs = 2,000
```

wellbeing	Proportion	Robust std. err.	z	P>|z|	[95% conf.	interval]
exercise (exercise vs control)	.7638805	.0154928	49.31	0.000	.7335151	.7942459

The indirect effect via bonotonin accounts for 76% of the effect of physical activity on well-being, and the remaining 24% is due to other mechanisms.

Example 2: Including covariates and relaxing the no-interaction assumption

The previous example was somewhat unrealistic. For causal inference, we must evaluate the potential of confounding. With causal mediation models, there are three types of confounders we should consider: treatment–outcome confounders, treatment–mediator confounders, and mediator–outcome confounders. A treatment–outcome confounder, for example, is a variable that affects both the selection into treatment and the outcome. If confounders exist and we observe them in our data, we can add them as covariates to the model to prevent biased results.

Above, we noted that the `wellbeing` data come from a randomized controlled trial. In this case, we do not have to worry about treatment–outcome and treatment–mediator confounders because treatment assignment is random. We do, however, need to consider variables such as `age`, `gender`, and `hstatus` (a person's health status) that affect both the mediator and the outcome. We include these variables as covariates in the model for well-being. We also make our model a bit more realistic by including baseline well-being in the outcome equation and baseline bonotonin level in the mediator equation. In addition, we omit the `nointeraction` option to allow the bonotonin coefficients to vary across treatment groups.

```
. mediate (wellbeing age gender i.hstatus basewell)
>         (bonotonin basebono)
>         (exercise)

Iteration 0:  EE criterion =  1.664e-25
Iteration 1:  EE criterion =  1.192e-28

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Linear
Mediator model:    Linear
Mediator variable: bonotonin
Treatment type:    Binary
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
exercise (Exercise vs Control)	9.941404	.2307909	43.08	0.000	9.489062	10.39375
NDE						
exercise (Exercise vs Control)	3.08372	.1684778	18.30	0.000	2.753509	3.41393
TE						
exercise (Exercise vs Control)	13.02512	.2356989	55.26	0.000	12.56316	13.48709

```
Note: Outcome equation includes treatment-mediator interaction.
```

The interpretation of the treatment effects is the same as before. The total effect of exercise on well-being is 13.0. Of this effect, 3.1 is attributed to the direct effect, while the remaining 9.9 is due to the indirect path via bonotonin. These results are similar to our simpler model above.

We find that the expected effect of exercise on well-being is 13.0, but what is the expected well-being when everyone exercises? When no one exercises? We can estimate four such potential-outcome means by specifying the pomeans option:

```
. mediate (wellbeing age gender i.hstatus basewell)
>         (bonotonin basebono)
>         (exercise), pomeans

Iteration 0:  EE criterion =  1.660e-25
Iteration 1:  EE criterion =  1.473e-28

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Linear
Mediator model:    Linear
Mediator variable: bonotonin
Treatment type:    Binary
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
Y0M0	56.94195	.2300492	247.52	0.000	56.49107	57.39284
Y1M0	60.02567	.2571311	233.44	0.000	59.52171	60.52964
Y0M1	66.78952	.2642177	252.78	0.000	66.27167	67.30738
Y1M1	69.96708	.232508	300.92	0.000	69.51137	70.42278

```
Note: Outcome equation includes treatment-mediator interaction.
```

Y1M1 is an estimate of the potential-outcome mean $E[Y_i(1, M_i(1))]$. If everyone in the population exercised, we would expect the average of well-being to be around 70. The values labeled Y1M0 and Y0M1 are estimates of the "cross-world" potential-outcome means $E[Y_i(1, M_i(0))]$ and $E[Y_i(0, M_i(1))]$. For these, we set different counterfactuals in the outcome and mediator equations. In this case, the Y1M0 estimate tells us the expected average well-being if, for the outcome equation, we assume that everyone in the population exercised, but we assume that no one exercised in regard to the effect of treatment on the mediator. If we compare the Y1M1 and Y1M0 estimates, we imagine a world where everyone received the treatment, except that the treatment is switched on and off in its effect on the mediator. The difference between these is $69.96708 - 60.02567 = 9.94141$, which is our NIE reported above.

By default, the TE, NIE, and NDE are computed, but we can request specific effects. For example, we could estimate only the NIE by typing

```
. mediate (wellbeing age gender i.hstatus basewell)
>         (bonotonin basebono)
>         (exercise), nie
```

Alternatively, we could estimate all available effects and potential-outcome means at once by specifying the `all` option:

```
. mediate (wellbeing age gender i.hstatus basewell)
>         (bonotonin basebono)
>         (exercise), all

Iteration 0:  EE criterion =  1.668e-25
Iteration 1:  EE criterion =  1.532e-28

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Linear
Mediator model:    Linear
Mediator variable: bonotonin
Treatment type:    Binary
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
Y0M0	56.94195	.2300492	247.52	0.000	56.49107	57.39284
Y1M0	60.02567	.2571311	233.44	0.000	59.52171	60.52964
Y0M1	66.78952	.2642177	252.78	0.000	66.27167	67.30738
Y1M1	69.96708	.232508	300.92	0.000	69.51137	70.42278
NIE						
exercise (Exercise vs Control)	9.941404	.2307909	43.08	0.000	9.489062	10.39375
NDE						
exercise (Exercise vs Control)	3.08372	.1684778	18.30	0.000	2.753509	3.41393
PNIE						
exercise (Exercise vs Control)	9.84757	.2318329	42.48	0.000	9.393186	10.30195
TNDE						
exercise (Exercise vs Control)	3.177554	.1800896	17.64	0.000	2.824585	3.530523
TE						
exercise (Exercise vs Control)	13.02512	.2356989	55.26	0.000	12.56316	13.48709

```
Note: Outcome equation includes treatment-mediator interaction.
```

Here we obtain estimates for two additional effects, PNIE and TNDE, which provide a different decomposition of the TE into direct and indirect effects. In this case, PNIE and TNDE are similar to NIE and NDE, respectively, because the coefficient on the treatment–mediator interaction term is quite small in the model for well-being. We can see the results for the underlying models, including this small coefficient of 0.002, if we add the `aequations` option to our `mediate` command.

Example 3: Referring to treatment effects using an alternative naming scheme

The effects we have discussed so far are sometimes referred to by different names. The default naming conventions originate in the works of Pearl and others. However, we can instead use terminology more closely tied to ATEs if we specify the `ateterms` option:

```
. mediate (wellbeing age gender i.hstatus basewell)
>         (bonotonin basebono)
>         (exercise), all ateterms
Iteration 0:  EE criterion =  1.668e-25
Iteration 1:  EE criterion =  1.532e-28
Causal mediation analysis                          Number of obs = 2,000
Outcome model:     Linear
Mediator model:    Linear
Mediator variable: bonotonin
Treatment type:    Binary
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
Y0M0	56.94195	.2300492	247.52	0.000	56.49107	57.39284
Y1M0	60.02567	.2571311	233.44	0.000	59.52171	60.52964
Y0M1	66.78952	.2642177	252.78	0.000	66.27167	67.30738
Y1M1	69.96708	.232508	300.92	0.000	69.51137	70.42278
AITE						
exercise (Exercise vs Control)	9.941404	.2307909	43.08	0.000	9.489062	10.39375
ADTE						
exercise (Exercise vs Control)	3.08372	.1684778	18.30	0.000	2.753509	3.41393
AITEC						
exercise (Exercise vs Control)	9.84757	.2318329	42.48	0.000	9.393186	10.30195
ADTET						
exercise (Exercise vs Control)	3.177554	.1800896	17.64	0.000	2.824585	3.530523
ATE						
exercise (Exercise vs Control)	13.02512	.2356989	55.26	0.000	12.56316	13.48709

Note: Outcome equation includes treatment-mediator interaction.

Using this notation, ATE can be decomposed into AITE and ADTE or into AITEC and ADTET. Notice that the estimates are the same as in the previous example; they now just have different names.

Example 4: Causal mediation model with a binary mediator

In the previous examples, both outcome and mediator variables were continuous. We now look at the case where the mediator variable is binary. To this end, we use the binary variable `bbonotonin`, an indicator of higher bonotonin levels after exercise, where improvement is defined as an increase of at least 10%. We could use a probit or a logit model for this mediator; we choose a logit model:

```
. mediate (wellbeing age gender i.hstatus basewell)
>         (bbonotonin, logit)
>         (exercise)

Iteration 0:  EE criterion =  8.253e-18
Iteration 1:  EE criterion =  8.223e-18  (backed up)

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Linear
Mediator model:    Logit
Mediator variable: bbonotonin
Treatment type:    Binary
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
exercise (Exercise vs Control)	4.41435	.4666635	9.46	0.000	3.499706	5.328994
NDE						
exercise (Exercise vs Control)	8.429238	.5696256	14.80	0.000	7.312792	9.545683
TE						
exercise (Exercise vs Control)	12.84359	.3712965	34.59	0.000	12.11586	13.57132

```
Note: Outcome equation includes treatment-mediator interaction.
```

Direct and indirect effect estimates differ from previous results because we used a different bonotonin measure as our mediator variable. However, because we still have the continuous well-being outcome, the interpretation of the effects is the same as before. Here we estimate a total effect of 12.8 with direct and indirect effects of 8.4 and 4.4, respectively. That is, we expect an increase of 12.8 in well-being due to treatment, of which 4.4 is due to an increase in bonotonin levels whereas the remaining 8.4 is due to other mechanisms.

Example 5: Causal mediation model with a binary outcome

Interpretation of effects did not change with a binary mediator, but interpretation does change when we specify a different type of outcome.

To demonstrate, we return to the continuous mediator but use a binary outcome variable. The outcome `bwellbeing` indicates higher well-being and is defined as an increase in well-being of at least 10% compared with the baseline measurement. Using `bwellbeing` as the outcome variable and specifying a probit outcome model, we get

```
. mediate (bwellbeing age gender i.hstatus, probit)
>         (bonotonin basebono, linear)
>         (exercise)

Iteration 0:  EE criterion =  2.177e-25
Iteration 1:  EE criterion =  9.730e-29

Causal mediation analysis                              Number of obs = 2,000

Outcome model:     Probit
Mediator model:    Linear
Mediator variable: bonotonin
Treatment type:    Binary
```

bwellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
exercise (Exercise vs Control)	.2346259	.0145763	16.10	0.000	.2060568	.263195
NDE						
exercise (Exercise vs Control)	.033732	.0237585	1.42	0.156	-.0128338	.0802978
TE						
exercise (Exercise vs Control)	.2683579	.0200872	13.36	0.000	.2289877	.3077281

```
Note: Outcome equation includes treatment-mediator interaction.
```

We interpret the effects as expected differences measured on the probability scale, sometimes referred to as risk differences. The TE of 0.27 indicates that if everyone in the population exercised, we would expect the probability of increased well-being to be 0.27 higher than the probability of increased well-being if no one exercised. In other words, the chance of experiencing an increase in well-being goes up by 27 percentage points when exposed to the exercise treatment. We can see that about 23 points are due to the indirect path via bonotonin, and about 3 points are due to other mechanisms.

Example 6: Causal mediation model with a binary mediator and binary outcome

We could also have the case where both the outcome and the mediator are binary. Here we use a logit model for both:

```
. mediate (bwellbeing age gender i.hstatus, logit)
>         (bbonotonin, logit)
>         (exercise)

Iteration 0:  EE criterion =  4.223e-16
Iteration 1:  EE criterion =  2.107e-30

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Logit
Mediator model:    Logit
Mediator variable: bbonotonin
Treatment type:    Binary
```

bwellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
exercise (Exercise vs Control)	.0959618	.0288699	3.32	0.001	.0393778	.1525457
NDE						
exercise (Exercise vs Control)	.1676141	.0358902	4.67	0.000	.0972706	.2379577
TE						
exercise (Exercise vs Control)	.2635759	.0212488	12.40	0.000	.221929	.3052228

```
Note: Outcome equation includes treatment-mediator interaction.
```

The interpretation is again in terms of differences in probabilities. We observe a TE of around 0.26, which is partially due to the indirect effect via `bbonotonin` (0.10) and partially due to other mechanisms (0.17).

Example 7: Causal mediation model with a count mediator

We use a fictional dataset on birthweights and demonstrate how to perform causal mediation analysis when using a Poisson model for a count mediator.

We now pretend to have observational data instead of experimental data. The sample includes women who gave birth to a child. We wish to find out whether socioeconomic status and education of the mother affects the child's health. The outcome variable is the birthweight of the baby (`bweight`), and the treatment variable is whether or not the mother has a college degree (`college`). The mediator variable is the number of cigarettes smoked per day during pregnancy (`ncigs`). The hypothesis is that women with a higher educational degree are likely to smoke fewer cigarettes and that smoking during pregnancy has negative effects on birthweight. Here is an excerpt from the dataset:

```
. use https://www.stata-press.com/data/r18/birthweight, clear
(Fictional birthweight data)
. list bweight ncigs college ses sespar age in 1/5
```

	bweight	ncigs	college	ses	sespar	age
1.	3621	1	No	5.3581	3.308523	29
2.	3278	0	Yes	9.556957	4.376035	38
3.	3073	1	No	3.980829	6.580275	39
4.	3306	0	Yes	11.17643	12.12075	30
5.	4517	0	Yes	9.026146	4.738766	28

We fit a linear model for the outcome `bweight` and a Poisson model for the mediator `ncigs`, and we specify `college` as the binary treatment variable. Because we have fully observational data, where selection into treatment is no longer completely random, we have to be concerned about all confounder types as mentioned in *Example 2: Including covariates and relaxing the no-interaction assumption*. We specify several potential confounders as covariates in both equations.

We do not assume that the adverse effects of smoking are different between women with a college degree and women without a college degree. Therefore, we use the `nointeraction` option.

```
. mediate (bweight sespar c.age##c.age)
>         (ncigs sespar c.age##c.age, poisson)
>         (college), nointeract
Iteration 0:  EE criterion =  1.939e-21
Iteration 1:  EE criterion =  1.937e-21  (backed up)
Causal mediation analysis                                   Number of obs = 2,000
Outcome model:     Linear
Mediator model:    Poisson
Mediator variable: ncigs
Treatment type:    Binary
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
college (Yes vs No)	167.3075	21.36134	7.83	0.000	125.4401	209.175
NDE						
college (Yes vs No)	347.3375	34.44561	10.08	0.000	279.8253	414.8496
TE						
college (Yes vs No)	514.645	28.65043	17.96	0.000	458.4912	570.7988

```
Note: Outcome equation does not include treatment-mediator interaction.
```

As before, the type of model we use for the mediator does not affect the interpretation of the estimated treatment effects. Effects are expected differences on the scale of the outcome variable. The TE indicates that if all women had a college degree, the average birthweight of newborn babies would be almost 515 grams higher than the average birthweight if no woman had a college degree. Of this weight increase, around 167 grams are due to women with higher educational degrees smoking less, while 347 grams are due to other mechanisms.

Example 8: Causal mediation model with an exponential-mean outcome

Here we use an exponential-mean model for the outcome `bweight`.

```
. mediate (bweight sespar c.age##c.age, expmean)
>         (ncigs sespar c.age##c.age, poisson)
>         (college), nointeract
Iteration 0:  EE criterion =  3.250e-13
Iteration 1:  EE criterion =  1.159e-17
Causal mediation analysis                              Number of obs = 2,000
Outcome model:     Exponential mean
Mediator model:    Poisson
Mediator variable: ncigs
Treatment type:    Binary
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
college (Yes vs No)	198.978	23.53279	8.46	0.000	152.8546	245.1014
NDE						
college (Yes vs No)	320.3318	34.47792	9.29	0.000	252.7563	387.9072
TE						
college (Yes vs No)	519.3098	28.70435	18.09	0.000	463.0503	575.5693

```
Note: Outcome equation does not include treatment-mediator interaction.
```

Because we are still modeling a continuous outcome, the interpretation does not change. The TE is about 519 grams, of which 199 grams are due to women with a college degree smoking less.

Example 9: Causal mediation model with multivalued treatment

So far we have only dealt with treatments that are binary. However, experiments often have more than two treatment arms, or an observational treatment could consist of multiple categories. Then we would refer to the treatment as multivalued.

To demonstrate, we return to our well-being data and use treatment variable `mexercise`, which captures three treatment groups: a control group, a group where individuals exercised for 45 minutes, and a group where individuals exercised for 90 minutes. Such a design would allow the researcher to find out whether and how the duration of exercise affects bonotonin levels and thereby well-being. Here we use a linear model for both the outcome and the mediator, and we include the multivalued treatment `mexercise` as our treatment variable:

```
. use https://www.stata-press.com/data/r18/wellbeing, clear
(Fictional well-being data)
. mediate (wellbeing age gender i.hstatus basewell)
>         (bonotonin basebono)
>         (mexercise)
Iteration 0:  EE criterion =  1.697e-25
Iteration 1:  EE criterion =  2.577e-26
Causal mediation analysis                          Number of obs = 2,000
Outcome model:     Linear
Mediator model:    Linear
Mediator variable: bonotonin
Treatment type:    Multivalued
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
mexercise (45 minutes vs Control)	5.128899	.3505171	14.63	0.000	4.441898	5.815899
mexercise (90 minutes vs Control)	9.780537	.2880877	33.95	0.000	9.215895	10.34518
NDE						
mexercise (45 minutes vs Control)	1.197498	.1750038	6.84	0.000	.8544965	1.540499
mexercise (90 minutes vs Control)	3.051084	.2071236	14.73	0.000	2.645129	3.457039
TE						
mexercise (45 minutes vs Control)	6.326396	.3894269	16.25	0.000	5.563134	7.089659
mexercise (90 minutes vs Control)	12.83162	.2967962	43.23	0.000	12.24991	13.41333

```
Note: Outcome equation includes treatment-mediator interaction.
```

We now have two effects per estimand because we compare the two treated groups to the control group. Starting with the TE, we expect nearly a 13-point increase in well-being if everyone in the population exercised for 90 minutes. Of these 13 points, around 10 points are due to the increase in bonotonin levels and 3 points are due to other mechanisms. The results for the 45-minute treatment arm, though expectedly smaller in magnitude, are interpreted similarly.

Example 10: Causal mediation model with continuous treatment

Instead of a binary or multivalued treatment, we could have a continuous treatment variable. With continuous treatments, we have to specify at least two values, one to be the treatment and another to be the control. We return to our birthweight data and use socioeconomic status (`ses`) as our continuous treatment variable. Here are some summary statistics for `ses`:

```
. use https://www.stata-press.com/data/r18/birthweight
(Fictional birthweight data)
. summarize ses
```

Variable	Obs	Mean	Std. dev.	Min	Max
ses	2,000	7.804412	2.287496	1.304026	16.27844

We can see that `ses` ranges from around 1 to 16 and has a mean of about 8. These values, however, do not tell us much because the variable is measured on an arbitrary scale. Therefore, we standardize it so that the resulting variable has a mean of 0 and a standard deviation of 1:

```
. generate std_ses = (ses-r(mean))/r(sd)
```

We will use the new variable, `std_ses`, as our treatment variable. We include the `continuous()` option within the third set of parentheses where we define the treatment. This option tells `mediate` to treat the variable as continuous and to use the values specified within the option as the control and treatment points. The first value is the control, and the remaining values are treatments that are compared with the control. Here we will specify one standard deviation below the mean as our control value and one standard deviation above the mean as our treatment value:

```
. mediate (bweight sespar c.age##c.age, expmean)
>         (ncigs sespar c.age##c.age, poisson)
>         (std_ses, continuous(-1 1)), nointeract

Iteration 0:  EE criterion =  1.470e-12
Iteration 1:  EE criterion =  1.980e-17

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Exponential mean
Mediator model:    Poisson
Mediator variable: ncigs
Treatment type:    Continuous

Continuous treatment levels:
  0: std_ses = -1 (control)
  1: std_ses =  1
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
std_ses (1 vs 0)	171.3015	14.68778	11.66	0.000	142.514	200.089
NDE						
std_ses (1 vs 0)	170.0598	32.14841	5.29	0.000	107.05	233.0695
TE						
std_ses (1 vs 0)	341.3613	31.73741	10.76	0.000	279.1571	403.5655

```
Note: Outcome equation does not include treatment-mediator interaction.
```

Even though we used a continuous treatment variable, we interpret the results as before: if everyone in the population had a socioeconomic status one standard deviation above the mean, the birthweight of newborn children would be about 341 grams higher than the birthweight if everyone's status value is one standard deviation below the mean. Of these 341 grams, roughly half is due to women with a higher status smoking less, and the other half is due to other mechanisms.

We could also evaluate the treatment effects at more than two values. Here we use the mean (0) of the standardized variable as the base, and we evaluate the treatment effects at −2, −1, 1, and 2:

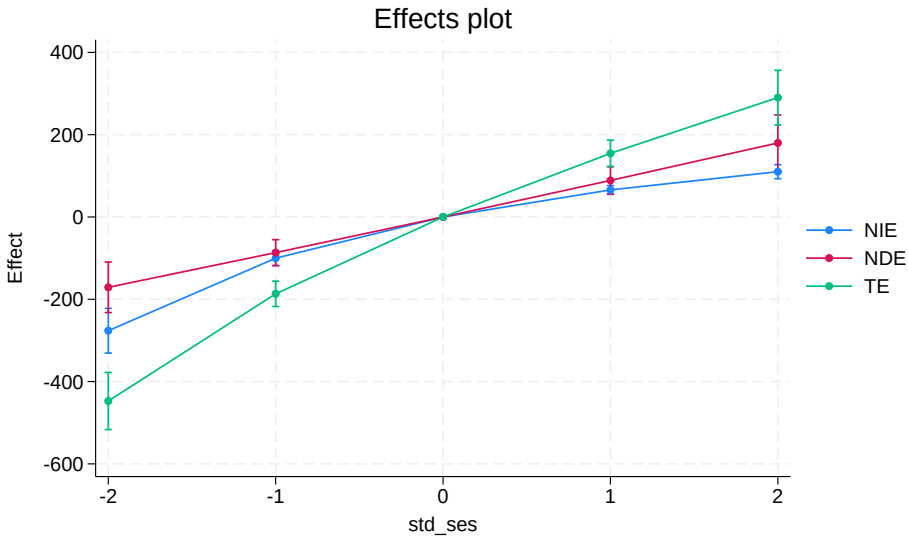
```
. mediate (bweight sespar c.age##c.age, expmean)
>         (ncigs sespar c.age##c.age, poisson)
>         (std_ses, continuous(0 -2 -1 1 2)), nointeract

Iteration 0:  EE criterion =  1.470e-12
Iteration 1:  EE criterion =  2.773e-17

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Exponential mean
Mediator model:    Poisson
Mediator variable: ncigs
Treatment type:    Continuous

Continuous treatment levels:
  0: std_ses =  0 (control)
  1: std_ses = -2
  2: std_ses = -1
  3: std_ses =  1
  4: std_ses =  2
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
std_ses						
(1 vs 0)	-276.2757	27.69004	-9.98	0.000	-330.5471	-222.0042
(2 vs 0)	-100.1155	9.170566	-10.92	0.000	-118.0894	-82.14148
(3 vs 0)	65.84585	5.423096	12.14	0.000	55.21678	76.47493
(4 vs 0)	110.1346	8.724232	12.62	0.000	93.03538	127.2337
NDE						
std_ses						
(1 vs 0)	-170.9012	31.33649	-5.45	0.000	-232.3196	-109.4828
(2 vs 0)	-86.56069	16.08129	-5.38	0.000	-118.0794	-55.04193
(3 vs 0)	88.83929	16.94031	5.24	0.000	55.6369	122.0417
(4 vs 0)	180.0172	34.77372	5.18	0.000	111.8619	248.1724
TE						
std_ses						
(1 vs 0)	-447.1769	35.41401	-12.63	0.000	-516.5871	-377.7667
(2 vs 0)	-186.6761	15.73291	-11.87	0.000	-217.5121	-155.8402
(3 vs 0)	154.6851	16.31969	9.48	0.000	122.6991	186.6712
(4 vs 0)	290.1517	33.85571	8.57	0.000	223.7958	356.5077

```
Note: Outcome equation does not include treatment-mediator interaction.
```

We now get four effect estimates for each treatment effect, which capture the expected differences in the outcome with respect to the control point. With multiple effect estimates, it can be convenient to plot the results. We use the postestimation command `estat effectsplot` to do so:

```
. estat effectsplot
```



For more information about `estat effectsplot`, see [CAUSAL] **mediate postestimation**.

Example 11: Estimating controlled direct effects

Controlled direct effects (CDEs) are different from the other estimands we have dealt with so far. Here, rather than having potential outcomes of the form $Y_i(t, M_i(t'))$, we have potential outcomes of the form $Y_i(t|M_i = m)$. That is, we have potential outcomes for each treatment level that are evaluated at set values of the mediator. Thus, CDEs only use the results of the outcome equation. Assuming a binary treatment, the CDE for value m of the mediator is $\text{CDE}(m) = Y_i(1|M_i = m) - Y_i(0|M_i = m)$. CDEs can be estimated using the postestimation command `estat cde`.

To demonstrate, we begin by fitting a mediation model using the well-being data:

```
. use https://www.stata-press.com/data/r18/wellbeing, clear
(Fictional well-being data)
. mediate (bwellbeing age gender i.hstatus, probit)
>         (bbonotonin, probit)
>         (exercise)
Iteration 0:  EE criterion =  4.326e-19
Iteration 1:  EE criterion =  5.424e-32
Causal mediation analysis                          Number of obs = 2,000
Outcome model:     Probit
Mediator model:    Probit
Mediator variable: bbonotonin
Treatment type:    Binary
```

bwellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
exercise (Exercise vs Control)	.0962832	.0288905	3.33	0.001	.0396588	.1529076
NDE						
exercise (Exercise vs Control)	.1672304	.0358936	4.66	0.000	.0968803	.2375805
TE						
exercise (Exercise vs Control)	.2635137	.0212346	12.41	0.000	.2218946	.3051327

```
Note: Outcome equation includes treatment-mediator interaction.
```

We fit probit models for the outcome and the mediator, but the type of model is not important here; we can use `estat cde` after any model fit with `mediate`.

What is the TE if everyone in the population has a mediator value of 0 (no improvement in bonotonin levels)? To find out, we estimate CDE(0) by specifying the `mvalue(0)` option with `estat cde`:

```
. estat cde, mvalue(0)
Controlled direct effect                           Number of obs = 2,000
Mediator variable: bbonotonin
Mediator value = 0
```

	CDE	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
exercise (Exercise vs Control)	.1605355	.039731	4.04	0.000	.0826641	.2384068

This CDE is around 0.16. The probability of increased well-being when everyone exercises is 0.16 higher than the probability of increased well-being when no one exercises, provided that no one in the population had at least a 10% increase in bonotonin levels.

We could perform the same analysis specifying multiple values for the mediator. Here we wish to estimate both CDE(0) and CDE(1):

```
. estat cde, mvalue(0 1)
Controlled direct effect                        Number of obs = 2,000
Mediator variable: bbonotonin
Mediator values:
  1._at: bbonotonin = 0
  2._at: bbonotonin = 1
```

	CDE	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
exercise@_at (Exercise vs Control) 1	.1605355	.039731	4.04	0.000	.0826641	.2384068
(Exercise vs Control) 2	.2224479	.0493025	4.51	0.000	.1258166	.3190791

If we "switch on" the mediator, the CDE is higher by around 0.06 points. We could also estimate this difference directly by using the `contrast` option:

```
. estat cde, mvalue(0 1) contrast
Controlled direct effect                        Number of obs = 2,000
Mediator variable: bbonotonin
Mediator values:
  1._at: bbonotonin = 0
  2._at: bbonotonin = 1
```

	CDE	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
_at#exercise (2 vs 1) (Exercise vs Control)	.0619124	.0630241	0.98	0.326	-.0616126	.1854373

See [CAUSAL] **mediate postestimation** for further information about `estat cde`.

Example 12: Estimating treatment effects on different scales

The `mediate` command estimates treatment effects on the natural scale of the outcome variable. However, some researchers may want to present their estimated effects on a different scale such as on the odds-ratio or risk-ratio scale if the outcome variable is binary or on the incidence-rate–ratio scale if the outcome variable is a count. The postestimation commands `estat rr`, `estat or`, and `estat irr` transform estimated treatment effects onto these different scales.

To see how this works, we first fit the following model with a binary outcome variable. We could use a probit or logit model for the outcome variable; here we are using `probit`:

```
. mediate (bwellbeing age gender i.hstatus, probit)
>         (bbonotonin, probit)
>         (exercise), all

Iteration 0:  EE criterion =  4.326e-19
Iteration 1:  EE criterion =  5.886e-32

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Probit
Mediator model:    Probit
Mediator variable: bbonotonin
Treatment type:    Binary
```

bwellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
Y0M0	.3014153	.0146737	20.54	0.000	.2726554	.3301752
Y1M0	.4686457	.0328309	14.27	0.000	.4042984	.5329931
Y0M1	.3536121	.0381972	9.26	0.000	.2787469	.4284773
Y1M1	.5649289	.0154435	36.58	0.000	.5346603	.5951976
NIE						
exercise (Exercise vs Control)	.0962832	.0288905	3.33	0.001	.0396588	.1529076
NDE						
exercise (Exercise vs Control)	.1672304	.0358936	4.66	0.000	.0968803	.2375805
PNIE						
exercise (Exercise vs Control)	.0521968	.0348642	1.50	0.134	-.0161357	.1205293
TNDE						
exercise (Exercise vs Control)	.2113169	.041136	5.14	0.000	.1306917	.291942
TE						
exercise (Exercise vs Control)	.2635137	.0212346	12.41	0.000	.2218946	.3051327

Note: Outcome equation includes treatment-mediator interaction.

Given that our outcome variable is binary and our outcome model is probit, the potential-outcome means are averaged probabilities, and because the treatment effects are differences between potential-outcome means, the estimated effects can be interpreted as risk differences. For example, the natural indirect effect is the difference between potential-outcome means Y1M1 and Y1M0:

```
. display _b[POmeans:Y1M1]-_b[POmeans:Y1M0]
.09628322
```

If, instead of interpreting the effect on the risk-difference scale, we wanted to interpret it on the risk-ratio scale, we could simply compute the ratio of the potential-outcome means:

```
. display _b[POmeans:Y1M1]/_b[POmeans:Y1M0]
1.2054499
```

This is what estat rr is doing:

```
. estat rr
Transformed treatment effects                         Number of obs = 2,000
```

bwellbeing	Risk ratio	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE exercise (Exercise vs Control)	1.20545	.0746555	3.02	0.003	1.06766	1.361023
NDE exercise (Exercise vs Control)	1.554817	.132328	5.19	0.000	1.315937	1.837062
PNIE exercise (Exercise vs Control)	1.173172	.1157431	1.62	0.105	.966905	1.423442
TNDE exercise (Exercise vs Control)	1.597595	.1778207	4.21	0.000	1.284469	1.987055
TE exercise (Exercise vs Control)	1.874254	.1043578	11.28	0.000	1.680482	2.09037

The total treatment effect is now decomposed into multiplicative components NIE and NDE as well as PNIE and TNDE. That is, taking their product, rather than their sum, will yield the total effect.

Similarly, we can express all effects on the odds-ratio scale by using `estat or`:

```
. estat or
Transformed treatment effects                          Number of obs = 2,000
```

bwellbeing	Odds ratio	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE exercise (Exercise vs Control)	1.472222	.1708025	3.33	0.001	1.172788	1.848106
NDE exercise (Exercise vs Control)	2.044157	.3042049	4.80	0.000	1.527008	2.736449
PNIE exercise (Exercise vs Control)	1.267908	.1933291	1.56	0.120	.9403671	1.709534
TNDE exercise (Exercise vs Control)	2.373558	.4231302	4.85	0.000	1.673622	3.366217
TE exercise (Exercise vs Control)	3.009452	.281479	11.78	0.000	2.505377	3.614945

Here the total average treatment effect is 3 on the odds-ratio scale and is composed of odds ratios 1.47 and 2.04 in regard to NIE and NDE, respectively, and of odds ratios 1.27 and 2.37 in regard to PNIE and TNDE. Typically, the treatment effects can be interpreted more intuitively on the risk-difference scale, but there may be applications where transforming them to the risk-ratio or odds-ratio scale is desirable.

Notice that `estat rr`, `estat or`, and `estat irr` require estimation of potential-outcome means with `mediate`. If the fitted model does not contain potential-outcome mean estimates, these `estat` commands will refit the model. The reestimation does not affect the results, but computation takes longer. See [CAUSAL] **mediate postestimation** for further information about `estat rr`, `estat or`, and `estat irr`.

Stored results

mediate stores the following in e():

Scalars

e(N)	number of observations
e(N_clust)	number of clusters
e(k_eq)	number of equations in e(b)
e(k_levels)	number of levels in treatment variable
e(rank)	rank of e(V)
e(interact)	1 if treatment–mediator interaction included, 0 otherwise
e(converged)	1 if converged, 0 otherwise

Macros

e(cmd)	mediate
e(cmdline)	command as typed
e(depvar)	name of outcome variable
e(mvar)	name of mediator variable
e(tvar)	name of treatment variable
e(omodel)	linear, logit, probit, poisson, or expmean
e(mmodel)	linear, logit, probit, poisson, or expmean
e(wtype)	weight type
e(wexp)	weight expression
e(title)	title in estimation output
e(clustvar)	name of cluster variable
e(tlevels)	levels of treatment variable
e(tvartype)	binary, multivalued, or continuous
e(control)	control level
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(properties)	b V
e(estat_cmd)	program used to implement estat
e(predict)	program used to implement predict
e(marginsnotok)	predictions disallowed by margins

Matrices

e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators

Functions

e(sample)	marks estimation sample

In addition to the above, the following is stored in r():

Matrices

r(table)	matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

Methods and formulas

`mediate` fits causal mediation models and estimates direct, indirect, and total treatment effects. Using the potential-outcomes framework, the estimated treatment effects are the result of contrasts between potential-outcome means. Without loss of generality, let T_i be a binary treatment, $t \in \{0, 1\}$, for observations $i = 1, \ldots, N$, and let Y_i be the outcome and M_i be the mediator variable. The potential-outcome means are

$$\text{POM}_{t,t'} \equiv E[Y_i(t, M_i(t'))]$$

The treatment effects are then defined as follows:

$$\begin{aligned}
\text{NIE} &\equiv E[Y_i(1, M_i(1)) - Y_i(1, M_i(0))] \\
\text{NDE} &\equiv E[Y_i(1, M_i(0)) - Y_i(0, M_i(0))] \\
\text{PNIE} &\equiv E[Y_i(0, M_i(1)) - Y_i(0, M_i(0))] \\
\text{TNDE} &\equiv E[Y_i(1, M_i(1)) - Y_i(0, M_i(1))] \\
\text{TE} &\equiv E[Y_i(1, M_i(1)) - Y_i(0, M_i(0))]
\end{aligned}$$

Synonyms for NIE, NDE, PNIE, TNDE, and TE are AITE, ADTE, AITEC, ADTET, and ATE, respectively.

The potential-outcome means are the result of an integral of the conditional expectation of the outcome with respect to the conditional distribution of the mediator (Imai, Keele, and Tingley 2010):

$$E[Y_i(t, M_i(t'))|X_i = x] = \int E[Y_i|M_i = m, T_i = t, \mathbf{X}_i = \mathbf{x}]\, dF[m|T_i = t', \mathbf{X}_i = \mathbf{x}] \tag{1}$$

They are estimated as the sample average

$$\widehat{\text{POM}}_{t,t'} = \frac{1}{N} \sum_{i=1}^{N} Y_i(t, M_i(t'))|\mathbf{X}_i = \mathbf{x}$$

The estimated treatment effects are then the result of differences between estimated potential-outcome means.

`mediate` uses analytical solutions for the integral in (1) for a variety of parametric outcome and mediator model combinations. Let $\mathbf{X}_i = \{\mathbf{W}_i, \mathbf{Z}_i\}$, the index function of the outcome model is

$$\eta_i^Y = \beta_0 + \beta_1 T_i + \beta_2 M_i + \beta_3 T_i M_i + \mathbf{W}_i \boldsymbol{\gamma} \tag{2}$$

and that of the mediator model is

$$\eta_i^M = \alpha_0 + \alpha_1 T_i + \mathbf{Z}_i \boldsymbol{\zeta} \tag{3}$$

where $\mathbf{W}_i$ and $\mathbf{Z}_i$ are potentially overlapping sets of covariates. If the `nointeraction` option is used, η_i^Y reduces to the simpler function where $\beta_3 = 0$. Depending on which model is specified, the expected values of the outcome and mediator follow these functional forms:

Model	Link function
linear	η_i
exponential mean	e^{η_i}
Poisson	e^{η_i}
logit	$\Pi(\eta_i)$
probit	$\Phi(\eta_i)$

Π and Φ are the cumulative logistic and cumulative normal distribution functions, respectively. Between the outcome and mediator models, all combinations of the above functional forms are allowed with the exception of logit outcome models in combination with linear or exponential-mean mediator models.

`mediate` uses the estimated coefficients from (2) and (3) to estimate $\widehat{\text{POM}}_{t,t'}$. Calculation of $\widehat{\text{POM}}_{t,t'}$ depends on the combination of functional forms of the outcome and mediator models. We define the following terms, where t represents counterfactual values for the treatment with respect to the outcome equation and t' represents treatment counterfactuals in regard to the mediator equation:

$$\nu_t = \beta_0 + \beta_1 t + \mathbf{W}_i\boldsymbol{\gamma}$$
$$\xi_{t'} = \alpha_0 + \alpha_1 t' + \mathbf{Z}_i\boldsymbol{\zeta}$$
$$\kappa_t = \beta_2 + \beta_3 t$$

If the outcome model is linear, we have

$$E[Y_i(t, M_i(t'))] = \nu_t + \Theta_m(\xi_{t'})\kappa_t$$

where $\Theta_m(\cdot)$ denotes the identity function if the mediator model is linear, the cumulative normal distribution function if probit, the cumulative logistic distribution function if logit, and the exponential function if exponential mean or Poisson. In this case, exponential mean and Poisson are synonyms when specifying the mediator model; notice, though, that this is not necessarily the case for other model combinations.

If the outcome model is probit and the mediator model is linear or exponential mean, we have

$$E[Y_i(t, M_i(t'))] = \Phi\left[\frac{\nu_t + \Gamma_m(\xi_{t'})\kappa_t}{\sqrt{1 + \kappa_t^2\sigma_m^2}}\right]$$

where $\Gamma_m(\cdot)$ denotes the identity function if the mediator model is linear and denotes the exponential function if it is exponential mean, and σ_m^2 is the error variance pertaining to the mediator model.

For probit and logit outcome models in combination with probit and logit mediator models, the potential outcomes are

$$E[Y_i(t, M_i(t'))] = \Lambda_y(\nu_t + \kappa_t)\Lambda_m(\xi_{t'}) + \Lambda_y(\nu_t)\{1 - \Lambda_m(\xi_{t'})\}$$

where $\Lambda_y(\cdot)$ is the cumulative normal distribution function if the outcome model is probit and the cumulative logistic distribution function if the outcome model is logit. $\Lambda_m(\cdot)$ denotes the cumulative normal distribution function if the mediator model is probit and the cumulative logistic distribution function if the mediator model is logit.

Regarding the outcome model, notice that, unlike the case of the mediator model, Poisson and exponential mean always refer to the same model. Thus, we can use the terms Poisson and exponential mean interchangeably in regard to the outcome model. The potential outcomes in the case of the exponential-mean outcome model and the linear or exponential-mean mediator model are

$$E[Y_i(t, M_i(t'))] = e^{\nu_t + \kappa_t\Gamma_m(\xi_{t'}) + (\kappa_t^2\sigma_m^2)/2}$$

where $\Gamma_m(\cdot)$ is the identity function if the mediator model is linear and is the exponential function if the mediator model is exponential mean. For probit and logit mediator models, the potential outcomes are

$$E[Y_i(t, M_i(t'))] = \Lambda_m(\xi_{t'})e^{\nu_t + \kappa_t} + \{1 - \Lambda_m(\xi_{t'})\}e^{\nu_t}$$

If the outcome model is exponential mean, probit, or logit, and if the mediator model is Poisson, the potential outcomes are

$$E[Y_i(t, M_i(t'))] = \sum_{j=0}^{K} \Psi_y(\nu_t + \kappa_t j) \frac{e^{j\xi_{t'}} e^{-e^{\xi_{t'}}}}{j!}$$

where j indexes the counts of the mediator variable, and $\Psi_y(\cdot)$ denotes the exponential function if the outcome model is exponential mean, denotes the cumulative normal distribution function if the outcome model is probit, and denotes the cumulative logistic distribution function if the outcome model is logit.

`mediate` uses a method of moments estimator, also known as an estimating equations estimator, to estimate all auxiliary and effect parameters as well as their variance–covariance matrix. For more information about the underlying `gmm` command, see [R] **gmm**.

The postestimation commands `estat or`, `estat rr`, and `estat irr` calculate treatment effects on different scales. If the outcome variable is binary and the model for the outcome variable is `logit` or `probit`, `estat or` computes marginal treatment effects on the odds-ratio scale, whereas `estat rr` computes marginal treatment effects on the risk-ratio scale. If the model for the outcome variable is `poisson` or `expmean`, `estat irr` computes marginal treatment effects on the incidence-rate–ratio scale. Let $Y_{tM_{t'}}$ be a shorthand for $E[Y_i(t, M_i(t'))]$; the treatment effects on risk-ratio and incidence-rate–ratio scales are ratios of potential-outcome means:

$$\begin{aligned}
\text{NIE}^{\text{RR}} &\equiv Y_{1M_1}/Y_{1M_0} \\
\text{NDE}^{\text{RR}} &\equiv Y_{1M_0}/Y_{0M_0} \\
\text{PNIE}^{\text{RR}} &\equiv Y_{0M_1}/Y_{0M_0} \\
\text{TNDE}^{\text{RR}} &\equiv Y_{1M_1}/Y_{0M_1} \\
\text{TE}^{\text{RR}} &\equiv Y_{1M_1}/Y_{0M_0}
\end{aligned}$$

For logit and probit outcome models, $Y_{tM_{t'}}$ are probabilities, and so the treatment effects on odds-ratio scale are

$$\begin{aligned}
\text{NIE}^{\text{OR}} &\equiv Y_{1M_1}/(1-Y_{1M_1})/\{Y_{1M_0}/(1-Y_{1M_0})\} \\
\text{NDE}^{\text{OR}} &\equiv Y_{1M_0}/(1-Y_{1M_0})/\{Y_{0M_0}/(1-Y_{0M_0})\} \\
\text{PNIE}^{\text{OR}} &\equiv Y_{0M_1}/(1-Y_{0M_1})/\{Y_{0M_0}/(1-Y_{0M_0})\} \\
\text{TNDE}^{\text{OR}} &\equiv Y_{1M_1}/(1-Y_{1M_1})/\{Y_{0M_1}/(1-Y_{0M_1})\} \\
\text{TE}^{\text{OR}} &\equiv Y_{1M_1}/(1-Y_{1M_1})/\{Y_{0M_0}/(1-Y_{0M_0})\}
\end{aligned}$$

Notice that, with all three of these scales, the total effect is the product of direct and indirect effects, rather than their sum.

CDEs use only the results of the outcome equation. Rather than having potential outcomes of the form $Y_i(t, M_i(t'))$, here we have potential outcomes $Y_i(t|M_i = m)$. That is, we have potential outcomes for each treatment level t that are evaluated at value m of the mediator. CDE(m) then is the average of the differences between potential outcomes. For binary treatment, CDE(m) is defined as $Y_i(1|M_i = m) - Y_i(0|M_i = m)$. Letting Y_{tm} be a shorthand for $Y_i(t|M_i = m)$, we have that

$$\begin{aligned}
\text{CDE(m)} &\equiv Y_{1m} - Y_{0m} \\
\text{CDE(m)}^{\text{RR}} &\equiv Y_{1m}/Y_{0m} \\
\text{CDE(m)}^{\text{IRR}} &\equiv Y_{1m}/Y_{0m} \\
\text{CDE(m)}^{\text{OR}} &\equiv Y_{1m}/(1-Y_{1m})/\{Y_{0m}/(1-Y_{0m})\}
\end{aligned}$$

Acknowledgments

Stata has an active research community adding features to the area of causal mediation. We would like to acknowledge their previous and ongoing contributions to the area: `paramed` by Hanhua Liu, Richard Emsley, Graham Dunn, Tyler VanderWeele, and Linda Valeri; `medeff` by Raymond Hicks and Dustin Tingley; `rwrmed` by Ariel Linden, Chuck Huber, and Geoffrey T. Wodtke; and many more. Type `search casual mediation` to see Stata's official and community-contributed features for causal mediation.

References

Baron, R. M., and D. A. Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51: 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173.

Holland, P. W. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81: 945–960. https://doi.org/10.2307/2289064.

Imai, K., L. Keele, and D. Tingley. 2010. A general approach to causal mediation analysis. *Psychological Methods* 15: 309–334. https://psycnet.apa.org/doi/10.1037/a0020761.

Imbens, G. W. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics* 86: 4–29. https://doi.org/10.1162/003465304323023651.

Nguyen, T. Q., I. Schmid, E. L. Ogburn, and E. A. Stuart. 2022. Clarifying causal mediation analysis: Effect identification via three assumptions and five potential outcomes. *Journal of Causal Inference* 10: 246–279. https://doi.org/10.1515/jci-2021-0049.

Nguyen, T. Q., I. Schmid, and E. A. Stuart. 2021. Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. *Psychological Methods* 26: 255–271. https://psycnet.apa.org/doi/10.1037/met0000299.

Pearl, J., and D. MacKenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

Robins, J. M., and S. Greenland. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3: 143–155. https://doi.org/10.1097/00001648-199203000-00013.

VanderWeele, T. J. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press.

Also see

[CAUSAL] **mediate postestimation** — Postestimation tools for mediate

[CAUSAL] **teffects** — Treatment-effects estimation for observational data

[CAUSAL] **teffects ra** — Regression adjustment

[SEM] **sem** — Structural equation model estimation command

[U] **20 Estimation and postestimation commands**