SAS/STAT® 14.3
User’s Guide
The CAUSALMED Procedure
Overview: CAUSALMED Procedure

The CAUSALMED procedure estimates causal mediation effects from observational data. In causal mediation analysis, there are four main variables of interest:

- an outcome variable $Y$
- a treatment variable $T$ that is hypothesized to have direct and indirect causal effects on the outcome variable $Y$ (in epidemiology, a treatment variable is also known as an exposure, denoted as $A$)
• a mediator variable $M$ that is hypothesized to be causally affected by the treatment variable $T$ and that itself has a direct effect on the outcome variable $Y$

• a set of pretreatment or background covariates that confound the observed relationships among $Y$, $T$, and $M$

The relationships among the first three variables represent the primary causal mediation effects of interest. The set of covariates represent background or pretreatment characteristics that confound the observed relationships among $Y$, $T$, and $M$. To focus on describing the causal mediation and related effects of interest, the roles of covariates are omitted for the present discussion and will be described later.

Figure 33.1 represents the first three main variables in a causal diagram (see Pearl 2009 for a general theory for causal diagrams).

In the terminology of causal diagrams, there are two causal pathways of effects from the treatment variable $T$:

• A direct pathway: $T \rightarrow Y$

• A mediated or indirect pathway: $T \rightarrow M \rightarrow Y$

For example, the analyses in “Getting Started: CAUSALMED Procedure” on page 2075 and Example 33.1 examine whether parental encouragement ($T$) affects children’s cognitive development ($Y$) and whether the effect of parental encouragement is mediated by children’s learning motivation ($M$). The analysis in Example 33.3 examines the effect of smoking ($T$) on infant mortality ($Y$) and its mediation effect through lowering birth weights ($M$).

The mediated or indirect pathway, $T \rightarrow M \rightarrow Y$, provides an explanation of how the cause $T$ leads to the outcome $Y$—that is, through a mechanism by which $T$ first affects a mediator $M$ and then affects $Y$ (see, for example, VanderWeele 2015). In structural equation modeling (see, for example, Bollen 1989), this pathway simply represents an indirect effect of $T$ on $Y$. Whereas the former interpretation emphasizes the role of the mediator in the mechanism, the latter interpretation emphasizes the root cause in treatment $T$. Correspondingly, there are two ways to interpret the direct pathway, $T \rightarrow Y$: as the causal effect of $T$ on $Y$ that is not through the mechanism involving $M$, or as the direct causal effect of $T$ on $Y$. As an analysis tool, the CAUSALMED procedure does not distinguish these alternative interpretations.

In traditional structural equation modeling, the diagram representation in Figure 33.1 is often expressed in the form of an additive linear model that excludes the interaction effect between $T$ and $M$ on $Y$, which is usually assumed to be continuous. As a result, the definitions and computation of mediation effects do not apply to situations in which interaction effects or nonlinear models are present.
Features of the CAUSALMED Procedure

However, under the counterfactual framework for causal mediation analysis (Robins and Greenland 1992; Pearl 2001), the same diagram representation in Figure 33.1 excludes neither the interaction nor nonlinear models. Basing on this counterfactual framework, the CAUSALMED procedure models the following variables in causal mediation analysis:

- outcome variable \( Y \): binary, continuous, or count
- treatment variable \( T \): binary or continuous
- mediator variable \( M \): binary or continuous
- covariates: categorical or continuous

Instead of linear models, PROC CAUSALMED supports a limited set of generalized linear models for describing the relationships among the \( Y, T, \) and \( M \) variables.

In practice, one of the biggest issues is the presence of confounding covariates in observational data. Confounding covariates are those pretreatment or background characteristics that are associated with the \( Y, T, \) and \( M \) variables before the treatment and mediation take place. They complicate the observed relationships among \( Y, T, \) and \( M \) by introducing extraneous associations that are not due to the direct and indirect pathways in the causal diagram for mediation analysis.

Therefore, for observational studies, statistical analysis that is based solely on specifying the causal diagram in Figure 33.1 does not lead to causal interpretation of the mediation and related effects. Various approaches have been proposed for dealing with confounding, including the weighting approach of Hong (2015) and the Monte Carlo simulation approach of Imai et al. (2010). The CAUSALMED procedure implements the regression approach of VanderWeele (2014).

The regression approach relies on correct specification of covariate effects in the outcome and mediator models for covariate effect adjustment. The CAUSALMED procedure enables you to specify these covariate effects in a causal mediation analysis.

Using the CAUSALMED procedure to establish causal interpretations of mediation and related effects requires an understanding of the terminology, concepts, and assumptions of causal mediation analysis. These technical aspects are explained in the section “Causal Mediation Effects: Definitions, Assumptions, and Identification” on page 2097.

Features of the CAUSALMED Procedure

The CAUSALMED procedure provides the following statements for specifying the regression models in a causal mediation analysis:

- The MODEL statement enables you to specify a generalized linear model for the outcome variable.
- The MEDIATOR statement enables you to specify a generalized linear model for the mediator variable.
- The COVAR statement enables you to specify the covariate effects.
Chapter 33: The CAUSALMED Procedure

The CAUSALMED procedure fits generalized linear models that have binary, negative binomial, Poisson, or normal distributions for the outcome and that have binary or normal distributions for the mediator. You can include interaction effects between the treatment and mediator variables and among the covariates.

If you specify covariate effects in the COVAR statement, the procedure incorporates them when it fits the outcome and mediator models. The model estimates are then used to compute various causal mediation effects.

PROC CAUSALMED computes the following main causal mediation effects:

- total effect (TE)
- controlled direct effect (CDE)
- natural direct effect (NDE)
- natural indirect effect (NIE)

For continuous outcomes, these effects are computed on the original continuous scale. For binary outcomes, these effects are computed on the odds ratio scale and excess relative risk scale. For definitions of these and other effects, see Valeri and VanderWeele (2013) and VanderWeele (2014).

In addition to computing the estimated effects, PROC CAUSALMED computes their percentages relative to the total effect, including the following measures:

- percentage mediated
- percentage due to interaction
- percentage eliminated

PROC CAUSALMED can compute various two-way, three-way, and four-way decompositions of the total effect and their percentages based on formulas of VanderWeele (2014). These decompositions provide more detailed analyses of the mediation and interaction effects.

PROC CAUSALMED computes a four-way decomposition that includes the following effects:

- controlled direct effect (CDE): the component that is due neither to interaction nor mediation
- reference interaction (IRF or INT_{ref}): the component that is due to interaction but not mediation
- mediated interaction (IMD or INT_{med}): the component that is due to both interaction and mediation
- pure indirect effect (PIE): the component that is due to mediation but not interaction

PROC CAUSALMED computes the following three-way decompositions:

- NDE + PIE + IMD: natural direct effect, pure indirect effect, and mediated interaction
- CDE + PIE + PAI: controlled direct effect, pure indirect effect, and portion attributed to interaction
PROC CAUSALMED computes the following two-way decompositions:

- **NDE + NIE**: natural direct effect and natural indirect effect
- **CDE + PE**: controlled direct effect and portion eliminated
- **TDE + PIE**: total direct effect and pure indirect effect

For more information about estimation of various causal mediation effects and total effect decompositions, see the section “Causal Mediation Effects: Definitions, Assumptions, and Identification” on page 2097.

Because causal mediation effects are generally defined conditionally on values of covariates, it is important that you interpret mediation effects at suitable levels of the covariates. You can specify the levels in the `EVALUATE` statement. If you do not specify the covariate levels, PROC CAUSALMED sets them at some default “averaged” levels, which are the same as in the default method that was implemented by Valeri and VanderWeele (2013).

Likewise, you can specify the treatment, control, and mediator levels in the `EVALUATE` statement to evaluate causal mediation effects that are dependent on the levels of these variables. For example, the controlled direct effect is always evaluated at a controlled level of the mediator variable. For more information about how to use the `EVALUATE` statement, see the section “Evaluating Causal Mediation Effects” on page 2104. For an illustration, see Example 33.2.

PROC CAUSALMED supports the computation of standard errors and confidence intervals for the effects by the following methods:

- analytic methods that are based on the asymptotic theory for maximum likelihood estimates
- bootstrap methods that are based on resampling

The default is the analytic method. You can request the bootstrap methods by specifying the `BOOTSTRAP` statement.

For more information about the maximum likelihood analytic approach, see the section “Causal Mediation Effects: Definitions, Assumptions, and Identification” on page 2097. For more information about bootstrapping, see the `BOOTSTRAP` statement and the section “Bootstrap Methods” on page 2111.

---

**Getting Started: CAUSALMED Procedure**

This section illustrates basic features of the CAUSALMED procedure for estimating total, direct, and indirect effects and their corresponding percentages.

The example presented in this section is patterned after the theoretical educational models that are discussed by Marjoribanks (1974). However, the data in this example are simulated, and neither the analysis nor the interpretation of procedure output mirrors that of Marjoribanks (1974).

A study is conducted to understand whether an encouraging environment provided by parents has an effect on the cognitive development of children. A key question is whether the effect of parental encouragement is due in part to its enhancement of children’s motivation to learn. Two pathways of parental encouragement effect are possible:
• a direct pathway, which can be denoted as \text{Encourage} \rightarrow \text{CogPerform}

• a mediated or indirect pathway, which can be denoted as \text{Encourage} \rightarrow \text{Motivation} \rightarrow \text{CogPerform}

In these pathways, the variable \text{Encourage} represents parental encouragement, the variable \text{Motivation} represents the learning motivation of the children, and the variable \text{CogPerform} represents the cognitive performance of the children. In the terminology of mediation analysis, \text{Encourage} is a treatment or an exposure, \text{Motivation} is a mediator, and \text{CogPerform} is an outcome.

A simulated sample of 300 observations is saved in a data set named Cognitive. Each observation has six variable values, as shown in Figure 33.2.

**Figure 33.2 Input Data Set**

<table>
<thead>
<tr>
<th>Obs</th>
<th>SubjectID</th>
<th>FamSize</th>
<th>SocStatus</th>
<th>Encourage</th>
<th>Motivation</th>
<th>CogPerform</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>7</td>
<td>31</td>
<td>36</td>
<td>40</td>
<td>103</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>27</td>
<td>36</td>
<td>40</td>
<td>103</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>25</td>
<td>35</td>
<td>40</td>
<td>99</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>6</td>
<td>29</td>
<td>36</td>
<td>40</td>
<td>103</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>4</td>
<td>22</td>
<td>33</td>
<td>37</td>
<td>79</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>2</td>
<td>23</td>
<td>34</td>
<td>38</td>
<td>87</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0</td>
<td>29</td>
<td>37</td>
<td>41</td>
<td>112</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>4</td>
<td>23</td>
<td>34</td>
<td>38</td>
<td>87</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>3</td>
<td>20</td>
<td>32</td>
<td>36</td>
<td>71</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>3</td>
<td>28</td>
<td>36</td>
<td>40</td>
<td>103</td>
</tr>
</tbody>
</table>

The variables are defined as follows:

• \text{CogPerform}: the child’s score on a cognitive test (outcome)

• \text{Encourage}: the sum score of the ratings of three items about parents’ encouraging behavior in a questionnaire (treatment)

• \text{FamSize}: the size of the child’s family

• \text{Motivation}: the sum score of the child’s levels of motivation as evaluated by the child, the teacher, and the primary caretaker (mediator)

• \text{SocStatus}: the child’s social status, which is an aggregate measure of household income, parents’ occupations, and parents’ educational levels

• \text{StudentID}: the child’s identifier

Variables \text{FamSize} and \text{SocStatus} are background or pretreatment characteristics that you would like to control for when observing various causal effects—either total, direct, or mediated.

First, consider an analysis in which the pretreatment characteristics are omitted. The following statements invoke PROC CAUSALMED to estimate various effects without controlling for background confounding variables:
proc causalmed data=Cognitive all;
  model CogPerform = Encourage Motivation;
  mediator Motivation = Encourage;
run;

The ALL option in the PROC CAUSALMED statement displays all available output. The MODEL statement specifies the outcome model for CogPerform, which is affected by Encourage and Motivation. The MEDIATOR statement specifies the mediator model for Motivation, which is affected only by Encourage.

The output produced by PROC CAUSALMED is displayed in Figure 33.3 through Figure 33.6.

Figure 33.3 echoes the modeling information and displays the number of observations read and used in the analysis; it also identifies the outcome, treatment, and mediator variables. By default, PROC CAUSALMED assumes normal distributions and identity links for the response variables in the outcome and mediator models because they are continuous.

**Figure 33.3 Model Information**

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Outcome Variable</td>
</tr>
<tr>
<td>Treatment Variable</td>
</tr>
<tr>
<td>Mediator Variable</td>
</tr>
<tr>
<td>Outcome Distribution</td>
</tr>
<tr>
<td>Outcome Link Function</td>
</tr>
<tr>
<td>Mediator Distribution</td>
</tr>
<tr>
<td>Mediator Link Function</td>
</tr>
</tbody>
</table>

Number of Observations Read 300
Number of Observations Used 300

Figure 33.4 presents the estimated effects. All effect estimates and percentage estimates are significant. The total effect estimate is 8.04, which is decomposed into the natural direct effect (NDE=4.28) and natural indirect effect (NIE=3.76). The estimated controlled direct effect (CDE) is 4.28, which is evaluated at the mean value of the mediator variable Motivation by default. In the current model, CDE is the same as NDE. The ‘Percentage Mediated’ is 46.74%. This means that slightly less than half of the parental encouragement effect on children’s cognitive development can be attributed to the enhancement of children’s learning motivation.

**Figure 33.4 Summary of Total, Direct, and Mediated Effects**

<table>
<thead>
<tr>
<th>Summary of Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect</td>
<td>8.0423</td>
<td>0.0320</td>
<td>7.9796 8.1050 251.30</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
<td>4.2835</td>
<td>0.1062</td>
<td>4.0754 4.4917 40.33</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
<td>4.2835</td>
<td>0.1062</td>
<td>4.0754 4.4917 40.33</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
<td>3.7588</td>
<td>0.1091</td>
<td>3.5449 3.9727 34.44</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>46.7377</td>
<td>1.3254</td>
<td>44.1400 49.3353 35.26</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>46.7377</td>
<td>1.3254</td>
<td>44.1400 49.3353 35.26</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Chapter 33: The CAUSALMED Procedure

The tables in Figure 33.5 and Figure 33.6 are useful for confirming the direction of the effects. Figure 33.5 shows the estimates of the outcome model for CogPerform.

**Figure 33.5** Estimates of the Outcome Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-201.2</td>
<td>0.6426</td>
<td>-202.5 -199.9</td>
<td>98053.6157</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Encourage</td>
<td>4.2835</td>
<td>0.1062</td>
<td>4.0754 4.4917</td>
<td>1626.7935</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Motivation</td>
<td>3.7576</td>
<td>0.1052</td>
<td>3.5514 3.9639</td>
<td>1274.6903</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Scale</td>
<td>0.4605</td>
<td>0.0188</td>
<td>0.4251 0.4989</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 33.6 shows the estimates of the mediator model for Motivation. The estimates of the direct effects from Encourage and Motivation are both positive and significant, thus confirming the positive effect of parental encouragement on children’s learning motivation.

**Figure 33.6** Estimates of the Mediator Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.0428</td>
<td>0.2641</td>
<td>3.5251 4.5605</td>
<td>234.2732</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Encourage</td>
<td>1.0003</td>
<td>0.0077</td>
<td>0.9853 1.0153</td>
<td>17040.9178</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Scale</td>
<td>0.2526</td>
<td>0.0103</td>
<td>0.2332 0.2737</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although the preceding analysis is interpretable, it does not take full advantage of the causal analytic techniques that are available in the CAUSALMED procedure. In order to draw valid causal interpretations from observational data, you must statistically control for all important confounding background characteristics.

Assume that FamSize and SocStatus are the only important confounding background characteristics that need to be controlled for. You can specify these variables as covariates in the COVAR statement and use PROC CAUSALMED as follows to fit an appropriate causal mediation model:

```plaintext
proc causalmed data=Cognitive;
  model CogPerform = Encourage Motivation;
  mediator Motivation = Encourage;
  covar FamSize SocStatus;
run;
```

When the confounding covariates FamSize and SocStatus are included, the procedure adjusts the estimates of the causal effects, leading to a new set of results which are summarized in Figure 33.7.
The total effect of Encourage on CogPerform is now 6.84, which is about 1 point lower than the total effect that is obtained without including the confounding covariates in the analysis (see Figure 33.4). This discrepancy suggests that parts of the observed association between Encourage and CogPerform are indeed due to their associations with the confounding background covariates. Failure to adjust for the confounding covariates led to inflated estimates of the total causal effect in Figure 33.4.

The natural direct effect (NDE) in the current analysis is 4.30, which is not much different from that of the preceding analysis. However, the natural indirect effect (NIE) is now 2.55, which is more than 1 point lower than the NIE in Figure 33.4. Finally, the ‘Percentage Mediated’ is now only 37%, which is almost 10% lower than the ‘Percentage Mediated’ (47%) that Figure 33.4 shows.

These results demonstrate that you must carefully consider the set of confounding covariates when conducting a causal mediation analysis. Causal analysis from observational data generally involves many assumptions that require serious attention. For instance, in the current example you could consider the following questions:

1. Why should the analysis assume that the variables Encourage and Motivation do not have an interaction effect on CogPerform? Is there any justification for this assumption?
2. If there is an interaction effect, what is the amount of this effect?
3. What justifies treating Encourage as the cause and CogPerform as the effect?
4. Is the causal sequence among the variables Encourage, Motivation, and CogPerform properly captured in the data?

You can address Questions 1 and 2 by fitting a more general model that includes the interaction term to determine whether the interaction effect is ignorable. PROC CAUSALMED supports outcome models that have interaction effects, as illustrated in Example 33.1, which presents a continuation of the current analysis.

Question 3 does not have a definite statistical answer. Substantive knowledge or existing evidence of the relationships is required to support these justifications.

An answer to Question 4 must also be justified by using substantive knowledge about the system of interest. In many systems there are temporal conditions that the data must satisfy so that the effects of the treatment on the outcome, the treatment on the mediator, and the mediator on the outcome can be observed.

Some researchers use longitudinal studies to establish the causal sequence. For instance, you can collect data in stages to ensure a proper temporal ordering of the causal, mediation, and outcome events. In this...
example, you could collect data for CogPerform several months after collecting data for Motivation, which were collected several years after you obtained the information about Encourage and any pretreatment confounders.

If you collect all the data at the same time point, you would need to justify that the parental encouragement pattern has long been established and that its effect on children’s learning motivation has been stabilized well before the children took the cognitive performance test. In addition, you would also need to justify that the background or pretreatment characteristics had been stabilized well before the measurements of the treatment, mediator, and the outcome. Substantive knowledge is required to support these justifications.

The role of CAUSALMED procedure is to estimate causal mediation effects given that all related assumptions are satisfied. The procedure can only serve as a tool to refute the presence of causal effects (when estimates are close to zero) given the model. The procedure cannot be used to establish causal interpretations of effects if the necessary methodological and statistical assumptions are not satisfied. For more information about assumptions of causal mediation analysis, see the section “Causal Mediation Effects: Definitions, Assumptions, and Identification” on page 2097.

Syntax: CAUSALMED Procedure

The following statements are available in the CAUSALMED procedure:

PROC CAUSALMED <options> ;
  CLASS variables <(options)> . . . <variable>(options)> </ global-options> ;
  MODEL outcome=effects </ model-options> ;
  MEDIATOR mediator=treatment ;
  COVAR effects ;
  EVALUATE <'label'> assignment <assignment . . . ></ options> ;
  BOOTSTRAP <options> ;
  FREQ variable ;
  BY variables ;

Together with the PROC CAUSALMED statement, the MODEL and MEDIATOR statements are essential to causal mediation analysis. The MODEL statement provides the model for the outcome variable; you use this statement to specify the effects of the treatment, the mediator, and possibly their interactions on the outcome variable. The MEDIATOR statement provides the model for the mediator variable; you use this statement to specify the effect of the treatment on the mediator variable.

In addition to the MODEL and MEDIATOR statements, you use the COVAR statement to specify the effects of confounding covariates. Because the assumption of no unmeasured confounding covariates is critical to the validity of causal mediation analysis for observational data, it is important that you specify all important covariate effects in this statement.

The CLASS statement, if provided, must precede the MODEL, MEDIATOR, and COVAR statements. The CLASS statement names the classification variables to be used in the analysis.

The following sections describe the PROC CAUSALMED statement and then describe the other statements in alphabetical order.
PROC CAUSALMED Statement

PROC CAUSALMED < options > ;

The PROC CAUSALMED statement invokes the CAUSALMED procedure. Table 33.1 summarizes the options available in the PROC CAUSALMED statement.

Table 33.1 Options Available in the PROC CAUSALMED Statement

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Set and Variable Options</strong></td>
<td></td>
</tr>
<tr>
<td>DATA=</td>
<td>Specifies the input SAS data set</td>
</tr>
<tr>
<td>DESCENDING</td>
<td>Reverses the order of levels of binary outcome and mediator variables</td>
</tr>
<tr>
<td>NAMELENGTH=</td>
<td>Specifies the length of effect names</td>
</tr>
<tr>
<td>ORDER=</td>
<td>Specifies the ordering method for the levels of the classification variable</td>
</tr>
<tr>
<td>RORDER=</td>
<td>Specifies the ordering method for the levels of the outcome variable</td>
</tr>
<tr>
<td><strong>Estimation and Analysis</strong></td>
<td></td>
</tr>
<tr>
<td>ALPHA=</td>
<td>Specifies the level for confidence intervals</td>
</tr>
<tr>
<td>CASECONTROL</td>
<td>Requests an analysis for a case-control study</td>
</tr>
<tr>
<td>DECOMP</td>
<td>Requests various decompositions of the total effect</td>
</tr>
<tr>
<td><strong>Displayed Output</strong></td>
<td></td>
</tr>
<tr>
<td>NOPRINT</td>
<td>Suppresses display of all output</td>
</tr>
<tr>
<td>PALL</td>
<td>Displays all output</td>
</tr>
<tr>
<td>PMEDMOD</td>
<td>Displays mediator model parameter estimates</td>
</tr>
<tr>
<td>POUTCOMEMOD</td>
<td>Displays outcome model parameter estimates</td>
</tr>
<tr>
<td>PSHORT</td>
<td>Displays only the basic modeling information and effects summary</td>
</tr>
<tr>
<td>PSUMMARY</td>
<td>Displays only the effects summary</td>
</tr>
<tr>
<td><strong>Technical Details</strong></td>
<td></td>
</tr>
<tr>
<td>NLOPTIONS</td>
<td>Specifies the optimization options for model fitting</td>
</tr>
<tr>
<td>SINGULAR=</td>
<td>Specifies the singularity criterion</td>
</tr>
<tr>
<td>THREADS=</td>
<td>Specifies the number of threads to use</td>
</tr>
</tbody>
</table>

You can specify the following options:

**ALPHA=p**

specifies the level $1 - p$ for constructing confidence intervals. By default, $p = 0.05$, which corresponds to $1 - p = 95\%$ confidence intervals. If $p$ is greater than 1, it is interpreted as a percentage and divided by 100. When multiple confidence intervals are constructed, this level is applied to each interval one at a time. This will not control the coverage probability of the intervals simultaneously. To control
familywise coverage probability, you might consider supplying a value of $p$ that is precomputed based on a method such as Bonferroni adjustment.

**CASECONTROL**
requests an analysis for a case-control study. When you specify this option, PROC CAUSALMED fits a mediator model by using only observations for subjects in the control group.

In case-control studies, a group of subjects is identified with a target outcome condition (for example, a disease). This group is called the case group. A second group, known as the control group, is formed by identifying subjects who are known to be absent of the target outcome, but whose background characteristics are the same as that of the case group. The values of the hypothesized exposure or treatment variable of the case and control groups are then compared to see whether the outcome can be attributed to the exposure or treatment variable.

**DATA=SAS-data-set**
specifies an input data set that contains the raw data. If the DATA= option is omitted, the most recently created SAS data set is used.

**DECOMP<=i>**
requests various decompositions of the total effect. By default, several two- and three-way decompositions and a four-way decomposition are computed. When you specify 2, 3, or 4 for $i$, decompositions up to an $i$-way decomposition are computed.

For continuous outcomes, the decomposition of the total effect is on the original continuous scale. For binary responses, the decomposition of the total effect is on the excess relative risk scale (VanderWeele 2014). In addition, PROC CAUSALMED displays the corresponding decompositions as percentages.

The four-way decomposition is described in the section “Causal Mediation Effects: Definitions, Assumptions, and Identification” on page 2097. It contains the following four components:

- CDE (controlled direct effect): the component effect that is not due to interaction or mediation
- IRF (reference interaction): the component effect that is due to interaction but not mediation (IRF is denoted as INT$_{ref}$ in VanderWeele (2014))
- IMD (mediated interaction): the component effect that is due to both interaction and mediation (IMD is denoted as INT$_{med}$ in VanderWeele (2014))
- PIE (pure indirect effect): the component effect that is due to mediation but not interaction

PROC CAUSALMED computes the following three-way decompositions:

- NDE + PIE + IMD: natural direct effect, pure indirect effect, and mediated interaction
- CDE + PIE + PAI: controlled direct effect, pure indirect effect, and portion attributed to interaction

PROC CAUSALMED computes the following two-way decompositions:

- NDE + NIE: natural direct effect and natural indirect effect
- CDE + PE: controlled direct effect and portion eliminated
- TDE + PIE: total direct effect and pure indirect effect

For more information about the logic and interpretations of these decompositions, see VanderWeele (2014) and VanderWeele (2015).
PROC CAUSALMED Statement

DESCENDING
DESCEND
DESC

sorts the levels of the binary outcome and the binary mediator variables in reverse of the specified order.

NAMELENGTH=n

specifies the maximum length of effect names in tables to be n characters, where n is a value between 20 and 128. By default, NAMELEN=20.

NLOPTIONS(nlo-options)

specifies options for the nonlinear optimization methods that are used for fitting the specified models. You can specify one or more of the following nlo-options separated by spaces:

ABSCONV=r
ABSTOL=r

specifies an absolute function convergence criterion by which minimization stops when \( f(\psi^{(k)}) \leq r \), where \( \psi \) is the vector of parameters in the optimization and \( f(\cdot) \) is the objective function. The default value of r is the negative square root of the largest double-precision value.

ABSFCONV=r

specifies an absolute function difference convergence criterion. Termination requires a small change of the function value in successive iterations,

\[ |f(\psi^{(k-1)}) - f(\psi^{(k)})| \leq r \]

where \( \psi \) denotes the vector of parameters that participate in the optimization and \( f(\cdot) \) is the objective function. By default, ABSFCONV=0.

ABSGCONV=r
ABSGTOL=r

specifies an absolute gradient convergence criterion. Termination requires the maximum absolute gradient element to be small,

\[ \max_j |g_j(\psi^{(k)})| \leq r \]

where \( \psi \) denotes the vector of parameters that participate in the optimization and \( g_j(\cdot) \) is the gradient of the objective function with respect to the \( j \)th parameter. By default, ABSGCONV=1E–7.

FCONV=r
FTOL=r

specifies a relative function convergence criterion. Termination requires a small relative change of the function value in successive iterations,

\[ \frac{|f(\psi^{(k)}) - f(\psi^{(k-1)})|}{|f(\psi^{(k-1)})|} \leq r \]

where \( \psi \) denotes the vector of parameters that participate in the optimization and \( f(\cdot) \) is the objective function. By default, FCONV=10^{–FDIGITS}, where by default FDIGITS is \( -\log_{10}\epsilon \), where \( \epsilon \) is the machine precision.
GCONV=r

GTOL=r

specifies a relative gradient convergence criterion. For all values of the TECHNIQUE= suboption except CONGRA, termination requires the normalized predicted function reduction to be small,

\[
g(k) \frac{|H(k)^{-1}g(k)|}{|f(k)|} \leq r
\]

where \( \psi \) denotes the vector of parameters that participate in the optimization, \( f(\cdot) \) is the objective function, and \( g(\cdot) \) is the gradient. When TECHNIQUE=CONGRA (for which a reliable Hessian estimate \( H \) is not available), the following criterion is used:

\[
\frac{\| g(\psi(k)) \|_2^2}{\| g(\psi(k)) - g(\psi(k-1)) \|_2 \| f(\psi(k)) \|} \leq r
\]

By default, GCONV=1E–8.

MAXFUNC=n

MAXFU=n

specifies the maximum number of function calls in the optimization process. The default values depend on the value of the TECHNIQUE= suboption as follows:

- TRUREG, NRRIDG, and NEWRAP: 125
- QUANEW and DBLDOG: 500
- CONGRA: 1,000

The optimization can terminate only after completing a full iteration. Therefore, the number of function calls that are actually performed can exceed \( n \).

MAXITER=n

MAXIT=n

specifies the maximum number of iterations in the optimization process. The default values depend on the value of the TECHNIQUE= suboption as follows:

- TRUREG, NRRIDG, and NEWRAP: 50
- QUANEW and DBLDOG: 200
- CONGRA: 400

These default values also apply when \( n \) is specified as a missing value.

MAXTIME=r

specifies an upper limit of \( r \) seconds of CPU time for the optimization process. Because the time is checked only at the end of each iteration, the actual run time might be longer than \( r \). By default, CPU time is not limited.
TECHNIQUE=CONGRA | DBLDOG | NEWRAP | NRRIDG | QUANEW | TRUREG

specifies the optimization technique to obtain maximum likelihood estimates. You can specify the following values:

CONGRA
performs a conjugate-gradient optimization.

DBLDOG
performs a version of double-dogleg optimization.

NEWRAP
performs a Newton-Raphson optimization that combines a line-search algorithm with ridging.

NRRIDG
performs a Newton-Raphson optimization with ridging.

QUANEW
performs a dual quasi-Newton optimization.

TRUREG
performs a trust-region optimization.

By default, TECHNIQUE=NRRIDG.

For more information about these optimization methods, see the section “Choosing an Optimization Algorithm” on page 508 in Chapter 19, “Shared Concepts and Topics.”

NOPRINT
suppresses all displayed output. For more information about the options for controlling output display, see the section “ODS Table Names” on page 2112.

ORDER=DATA | FORMATTED | FREQ | INTERNAL

specifies the sort order for the levels of CLASS variables. This ordering determines which parameters in the model correspond to each level in the data.

You can specify the following values:

DATA
sorts the levels in their order of appearance in the input data set.

FORMATTED
sorts the levels by external formatted values, except for numeric variables that have no explicit format, which are sorted by their unformatted (internal) values. The sort order is machine-dependent.

FREQ
sorts the levels by descending frequency count. Levels that have more observations come earlier in the order.

INTERNAL
sorts the levels by an unformatted value. The sort order is machine-dependent.

By default, ORDER=FORMATTED. For more information about sort order, see the chapter on the SORT procedure in the SAS Visual Data Management and Utility Procedures Guide and the discussion of BY-group processing in SAS Language Reference: Concepts.

PALL

ALL

displays all output tables. For more information about the options for controlling output display, see the section “ODS Table Names” on page 2112.
PMEDMOD
displays parameter estimates for the mediator model. For more information about the options for controlling output display, see the section “ODS Table Names” on page 2112.

POUTCOMEMOD
displays parameter estimates for the outcome model. For more information about the options for controlling output display, see the section “ODS Table Names” on page 2112.

PSHORT
displays only the basic modeling information and the summary of effects. When you specify this option, you can also display the effect decomposition table by specifying the DECOMP option. For more information about the options for controlling output display, see the section “ODS Table Names” on page 2112.

PSUMMARY
displays only a summary of effects. When you specify this option, you can also display the effect decomposition table by specifying the DECOMP option. For more information about the options for controlling output display, see the section “ODS Table Names” on page 2112.

RORDER=DATA | FORMATTED | FREQ | INTERNAL
RESPORDER=DATA | FORMATTED | FREQ | INTERNAL
specifies the sort order for the levels of the outcome variable. In order for this option to apply, either the outcome variable must be specified in the CLASS statement or the DIST=BIN option must be specified in the MODEL statement. The following table shows how PROC CAUSALMED interprets values of the RORDER= option.

<table>
<thead>
<tr>
<th>Value of RORDER=</th>
<th>Levels Sorted By</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA</td>
<td>Order of appearance in the input data set.</td>
</tr>
<tr>
<td>FORMATTED</td>
<td>External formatted value, except for numeric variables that have no explicit format, which are sorted by their unformatted (internal) value. The sort order is machine-dependent.</td>
</tr>
<tr>
<td>FREQ</td>
<td>Descending frequency count. Levels that have the most observations come first in the order.</td>
</tr>
<tr>
<td>INTERNAL</td>
<td>Unformatted value. The sort order is machine-dependent.</td>
</tr>
</tbody>
</table>

By default, RORDER=FORMATTED. The DESCENDING option in the PROC CAUSALMED statement causes the response variable to be sorted in reverse of the order displayed in the previous table. For more information about sort order, see the chapter on the SORT procedure in the Base SAS Procedures Guide.

SINGULAR=tolerance
specifies the tolerance for testing the singularity of a matrix, where tolerance must be between 0 and 1. The default tolerance is 1E7 times the machine precision.
THREADS=n
NTHREADS=n

specifies the number of threads (n) for analytic computations and overrides the SAS system option THREADS | NOTHREADS. If you do not specify the THREADS= option or if you specify THREADS=0, the number of threads is determined from the number of CPUs in the host on which the analytic computations execute.

**BOOTSTRAP Statement**

```plaintext
BOOTSTRAP < options > ;
```

The BOOTSTRAP statement requests bootstrap estimates of standard errors and bootstrap confidence intervals for various effects and percentages of total effects.

Table 33.2 summarizes the options available in the BOOTSTRAP statement.

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOOTCI</td>
<td>Produces bootstrap confidence intervals for effects and percentages</td>
</tr>
<tr>
<td>NBOOT=</td>
<td>Specifies the number of bootstrap sample data sets (replicates)</td>
</tr>
<tr>
<td>SEED=</td>
<td>Specifies the seed that initializes the random number stream</td>
</tr>
</tbody>
</table>

You can specify the following options:

**BOOTCI < (BC | NORMAL | PERC | ALL)>**

**CI < (BC | NORMAL | PERC | ALL)>**

computes bootstrap-based confidence intervals for the effects and percentages of effects, and displays them in the “Summary of Effects” table. This table includes a column that indicates the number of bootstrap samples used to compute the confidence intervals; this column is not displayed but is available if you save the table as an output data set by using the ODS OUTPUT statement. You can also display this column by modifying the corresponding template.

You can specify one or more of the following types of bootstrap confidence intervals separated by spaces:

**ALL** produces all three confidence intervals, which are described in the following types.

**BC** produces bias-corrected confidence intervals. You must specify a value of 1,000 or more for the NBOOT= option, but the confidence intervals are not computed if fewer than 900 bootstrap replicates produce bootstrap estimates.

**NORMAL** produces confidence intervals that are based on the assumption that bootstrap estimates follow a normal distribution. You must specify a value of 50 or more for the value NBOOT= option, but the corresponding standard errors and confidence intervals are not computed if fewer than 40 bootstrap replicates produce bootstrap estimates.
**PERC** produces percentile-based confidence intervals. You must specify a value of 1,000 or more for the NBOOT= option, but the confidence intervals are not computed if fewer than 900 bootstrap replicates produce bootstrap estimates.

The **ALPHA=** option in the **PROC CAUSALMED** statement sets the confidence level for constructing bootstrap intervals. For more information about how bootstrap-based confidence intervals are computed, see the section “Bootstrap Methods” on page 2111. By default, **PROC CAUSALMED** produces bias-corrected confidence intervals (BOOTCI(BC)) based on 1,000 bootstrap samples.

**NBOOT=** \(n\)

**NSAMPLE=** \(n\)

**NSAMPLES=** \(n\)

specifies the number of bootstrap sample data sets (replicates), where \(n\) is between 50 and 10,000. By default, \(n=1000\).

**SEED=** \(n\)

provides the seed that initializes the random number stream for generating the bootstrap sample data sets (replicates). If you do not specify this option or if you specify a value for \(n\) that is less than or equal to 0, the seed is generated from reading the time of day from the computer’s clock. The largest possible value for the seed is \(2^{31} - 1\).

You can use the **SYSRANDOM** and **SYSRANEND** macro variables after a **PROC CAUSALMED** step to query the initial and final seed values. However, using the final seed value as the starting seed for a subsequent analysis does not continue the random number stream where the previous analysis ended. The **SYSRANEND** macro variable provides a mechanism to pass on seed values to ensure that the sequence of random numbers is the same every time you run an entire program. To reproduce the random number stream that was used to generate bootstrap estimates, you must specify the same value for the **SEED=** option and the same value for the **THREADS=** option in the **PROC CAUSALMED** statement.

**BY Statement**

**BY** \(variables\);

You can specify a **BY** statement with **PROC CAUSALMED** to obtain separate analyses of observations in groups that are defined by the **BY** variables. When a **BY** statement appears, the procedure expects the input data set to be sorted in order of the **BY** variables. If you specify more than one **BY** statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the **SORT** procedure with a similar **BY** statement.

- Specify the **NOTSORTED** or **DESCENDING** option in the **BY** statement for the **CAUSALMED** procedure. The **NOTSORTED** option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the **BY** variables) and that these groups are not necessarily in alphabetical or increasing numeric order.

- Create an index on the **BY** variables by using the **DATASETS** procedure (in Base SAS software).
For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *SAS Visual Data Management and Utility Procedures Guide*.

---

**CLASS Statement**

```
CLASS variable <(options)> . . . <variable <(options)>> < / global-options> ;
```

The CLASS statement names one or more classification variables to be used as explanatory variables in the analysis.

The CLASS statement must precede the COVAR, MEDIATOR, and MODEL statements. Most options can be specified either as individual variable options or as global-options. You can specify options for each variable by enclosing the options in parentheses after the variable name. You can also specify global-options for the CLASS statement by placing them after a slash (/). Global-options are applied to all the variables specified in the CLASS statement. However, individual CLASS variable options override the global-options. Unless otherwise indicated, you can specify the following values for either an option or a global-option:

**CPREFIX=n**

uses at most the first \( n \) characters of a CLASS variable name in creating names for the corresponding design variables. The default is \( 32 - \min(32, \max(2, f)) \), where \( f \) is the formatted length of the CLASS variable.

**DESCENDING**

**DESC**

reverses the sort order of the CLASS variables. If both the DESCENDING and ORDER= options are specified, PROC CAUSALMED orders the categories according to the ORDER= option and then reverses that order.

**LPREFIX=n**

uses at most the first \( n \) characters of a CLASS variable name in creating labels for the corresponding design variables. The default is \( 256 - \min(256, \max(2, f)) \), where \( f \) is the formatted length of the CLASS variable.

**MISSING**

treats missing values (blanks for character variables and ., ._, .A, . . . , .Z for numeric variables) as valid values for the CLASS variables.

**ORDER=DATA | FORMATTED | FREQ | INTERNAL**

specifies the sort order for the levels of CLASS variables. This ordering determines which parameters in the model correspond to each level in the data.

You can specify the following values:

**DATA**

sorts the levels in their order of appearance in the input data set.

**FORMATTED**

sorts the levels by external formatted values, except for numeric variables that have no explicit format, which are sorted by their unformatted (internal) values. The sort order is machine-dependent.
FREQ sorts the levels by descending frequency count. Levels that have more observations come earlier in the order.

INTERNAL sorts the levels by an unformatted value. The sort order is machine-dependent.

By default, ORDER=FORMATTED. For more information about sort order, see the chapter on the SORT procedure in the SAS Visual Data Management and Utility Procedures Guide and the discussion of BY-group processing in SAS Language Reference: Concepts.

**REF='level' | FIRST | LAST**

specifies a level of the CLASS variable to be put at the end of the list of levels. This level thus corresponds to the reference level in the usual interpretation of the linear estimates that have a singular parameterization.

You can specify the following values:

'level' specifies the level of the variable to use as the reference level. Specify the formatted value of the variable if a format is assigned. You cannot specify 'level' as a global-option.

FIRST designates the first ordered level as the reference level.

LAST designates the last ordered level as the reference level.

By default, REF=LAST.

**TRUNCATE< =n>**

specifies the length (n) of variable values to use in determining the CLASS variable levels. The default is to use the full formatted length of the CLASS variable. If you specify this option without the length n, the first 16 characters of the formatted values are used. When formatted values are longer than 16 characters, you can use this option to revert to the levels as determined in releases before SAS 9. The TRUNCATE option is available only as a global-option.

---

**COVAR Statement**

**COVAR effects ;**

The COVAR statement specifies the effects of covariates in a causal mediation analysis. These covariates represent important confounders in the causal model. You do not need to distinguish confounders for the treatment-outcome, treatment-mediator, or mediator-outcome relationships. You simply enter all confounding covariate effects in this statement. PROC CAUSALMED appends these effects into the design matrices of the outcome and mediator models that you specify in the MODEL and MEDIATOR statements, respectively. The causal mediation and other related effects are thus estimated with adjustment for confounding by including covariate effects in outcome and mediator modeling.

The simplest form of effects is a list of confounding covariates. For example, the following statement specifies that C1, C2, and C3 are confounding covariates in the causal model:

```
covar C1 C2 C3;
```

You can also include interaction terms in the specification. For example, the following statement adds the interaction of C1 and C2 as a confounding effect to the preceding specification:
EVALUATE Statement

EVALUATE < 'label' > assignment < assignment . . . > </ options > ;

You use the EVALUATE statement to specify variable levels or values for evaluating various effects. In the assignments, you can specify one or more of the following variable levels:

- the control and treatment levels for computing all effects
- the mediator level for computing the controlled direct effect
- the covariate levels for evaluating various conditional causal effects

Each assignment is of the form

\[ \text{var-key} = \text{value-key} \]

where var-key specifies a variable and value-key is either its numerical value, its character value, or a keyword (such as MEAN) that generates a value from the variable.

Because of interaction effects and nonlinear models, computation of causal mediation effects usually depend on the values or levels of the treatment, control, mediator, or covariate levels. PROC CAUSALMED assigns values or levels to these variables automatically when it performs a default mediation effect analysis. By setting these default variable levels, you obtain “overall” measures of causal mediation and related effects. For more information about how PROC CAUSALMED sets the default values of levels, see the section “Evaluating Causal Mediation Effects” on page 2104.

However, to address your particular research questions more directly, you can provide EVALUATE statements with specific variable levels to evaluate mediation and related effects. Specifying covariate levels, the treatment level (of the treatment variable), or the control level (of the treatment variable) changes the estimates of all mediation effects and decompositions. Specifying the controlled level of the mediator variable does not change the estimates of the total effect (TE), the natural direct effect (NDE), or the natural indirect effect (NIE). But it does change the estimates of the controlled direct effect (CDE) and the reference interaction (IRF).

You can provide as many EVALUATE statements as you want. Each statement specifies an assignment scheme that defines the mediation effects and produces a summary of effects, decompositions of effects (if requested), and percentage decompositions of effects (if requested).
To distinguish the results that are produced by different EVALUATE statements, you can specify a distinct label in each EVALUATE statement. A maximum of 256 characters is allowed for each label. This label is displayed in the output tables.

For example, suppose that C1 and C2 are continuous covariates in the mediation model and you want to evaluate the mediation effects at C1=5 and C2=10. You can request that by providing the following statement:

    evaluate 'Set C1=5 C2=10' C1=5 C2=10;

In this statement, the quoted string, 'Set C1=5 C2=10', labels the set of assignments for evaluating the mediation effects and is followed by the assignments, C1=5 and C2=10.

If you want to evaluate the mediation effects conditioned on a different set of covariate values, you can add another EVALUATE statement. For example,

    evaluate 'Scheme 1 -- C1=5 C2=10' C1=5 C2=10;
    evaluate 'Scheme 2 -- C1=10 C2=5' C1=10 C2=5;

Meaningful labels for the EVALUATE statements are highly recommended in practice.

If you use '_Default' as the label, PROC CAUSALMED overrides the default variable levels for evaluating mediation effects. For example, the following statement generates only one set of mediation effect output tables, which replace the default tables:

    evaluate '_Default' C1=5 C2=10; /* Overrides the default assignment scheme */

In addition to the use of fixed value assignments (such as C1=5 in the preceding examples), PROC CAUSALMED provides several ways to specify the var-key and the value-key in an assignment.

You can use the following var-keys in an assignment:

**CONTROL | A0 | T0**

varname(CONTROL) specifies the control level of the treatment variable, where varname represents the actual treatment variable name.

covariate-name specifies a covariate by using its actual variable name for covariate-name.

**MEDIATOR | MSTAR**

varname specifies the controlled level of the mediator variable, where varname represents the actual mediator variable name.

**TREATMENT | A1 | T1**

varname(TREATMENT) specifies the treatment level of the treatment variable, where varname represents the actual treatment variable name.

In all the preceding examples, actual variable names have served as var-keys and numerical values have served as value-keys. The following statements show examples of assignments that specify keywords for var-keys:
evaluate 'Scheme 3' _treatment=max _control=mean _mediator=last C1=mode;
evaluate 'Scheme 4' _A1=0 _A0=1 _mstar=10 C1='Boys';
evaluate 'Scheme 5' _A1=.5 _A0=-.5 _mediator=0 C1=2;

You can use the following value-keys in an assignment:

'value'<(SD)>

assigns the numerical value in the assignment, where value represents a fixed number. If you use the SD option, the measurement scale of the numerical value refers to the measurement scale of the standardized variable. Hence, with the SD option the actual assigned value is

\[ mean + value \times sd \]
where mean and sd are the sample mean and standard deviation, respectively, of the corresponding variable that is specified in the var-key of the assignment.

The following statements show examples of assignments that use different types of keywords for value-keys:

```
evaluate 'Evaluation 6' _treatment=.5(SD) _control=-0.5(SD) _mediator=min C1=mode;
evaluate 'Evaluation 7' _A1=first _A0=last _mediator=mean C1='Boys';
```

After specifying the assignments in an EVALUATE statement, you can use one or more of the following options to control the displays of mediation effects and decompositions that are generated by the EVALUATE statement:

- **CLABEL='clabel'**
  - specifies a short content-label for the mediation effects that are generated by the EVALUATE statement. The clabel is used to label the corresponding set of tables in the output. Only the first 20 characters of clabel is used. If you do not specify this option, the first 20 characters of the label, which you specify in the beginning of the EVALUATE statement, is used as clabel.

- **DECOMP<i>/</i>**
  - specifies the type of decompositions requested, where i is between 2 and 4, representing two-, three-, or four-way decompositions, respectively. If you specify the DECOMP= options in the PROC CAUSALMED statement and in an EVALUATE statement, the DECOMP option in the EVALUATE statement is used for evaluating the requested effects.

- **NODECOMP**
  - suppresses the display of all decomposition results for the specified evaluation scheme of mediation effects. Only the summary table of effects is shown. This option also overrides the DECOMP= option in the same EVALUATE statement.

For more information about how variable levels are related to the interpretation of causal mediation effects, see the section “Evaluating Causal Mediation Effects” on page 2104. For illustrations of the use of the EVALUATE statement, see Example 33.2 and Example 33.3.

---

**FREQ Statement**

```
FREQ variable;
```

If a variable in the data set represents the frequency of occurrence for the other values in the observation, include the variable’s name in a FREQ statement. The procedure then treats the data set as if each observation appears \( n \) times, where \( n \) is the value of the FREQ variable for the observation. The total number of observations is considered to be equal to the sum of the FREQ variable values when the procedure determines degrees of freedom for significance probabilities.

If the value of the FREQ variable is missing or is less than 1, the observation is not used in the analysis. If the value is not an integer, the value is truncated to an integer.
MEDIATOR Statement

MEDIATOR mediator=treatment ;

The MEDIATOR statement is required for specifying the mediator model. You provide mediator (the name of the mediator variable) on the left of the equal sign and treatment (the name of the treatment variable) on the right. For example, the following statement specifies M as the mediator variable and T as the treatment variable in the analysis:

mediator M = T;

Together, the COV AR, MEDIATOR, and MODEL statements specify the relationships of all variables in the mediation analysis. The mediator and treatment variables that you specify in the MEDIATOR statement must be consistent with those that you specify in the MODEL statement. If there are covariates in the analysis, do not specify them or their effects in the MEDIATOR statement even though covariate effects on the mediator variable are being modeled. Instead, use the COV AR statement to specify covariate effects.

Mediator and treatment variables can either be binary or continuous. You can specify that a mediator or treatment variable is a binary variable by listing it in the CLASS statement. Otherwise, it is assumed to be continuous.

PROC CAUSALMED assumes a normal distribution and the identity link function for a continuous mediator, and a binary distribution and the logit link function for a binary mediator.

MODEL Statement

MODEL outcome=effects < / model-options> ;

The MODEL statement is required for specifying the outcome model. You provide outcome (the name of the outcome variable) on the left of the equal sign and effects (the treatment and mediator effects) on the right.

Together, the COV AR, MEDIATOR, and MODEL statements specify the relationships of all variables in the mediation analysis. The treatment and mediator variables that you specify in the MODEL statement must be consistent with those that are you specify in the MEDIATOR statement. If there are covariates in the analysis, do not specify them or their effects in the MODEL statement even though the covariate effects on the outcome variable are being modeled. Instead, use the COV AR statement to specify the covariate effects.

Outcome variables can be binary, continuous, or count variables. PROC CAUSALMED does not support outcome variables that are nominal or ordinal and have more than two levels.

You can specify that an outcome variable is a binary variable by listing it in the CLASS statement. Alternatively, you can specify that an outcome variable is a binary variable by specifying DIST=BIN as a model-option in the MODEL statement. Outcome variables that are not specified as binary variables are treated as continuous variables.

Suppose that the outcome variable is Y, the treatment variable is T, and the mediator variable is M. The three possibilities for the syntax of effects are as follows:
model Y = T M;
model Y = T M T*M;
model Y = T | M;

The first statement specifies the effects of T and M but no interaction effect between the two. The second and third statements are equivalent. Both specify the effects of T, M, and their interaction. The order of T and M is not important.

You can specify the distribution and link function for the outcome model by providing the following `model-options` after the slash (/):

**DIST=** *keyword*

**DISTRIBUTION=** *keyword*

specifies the built-in probability distribution to use in the model. If you specify this option and you omit the `LINK=` option, a default link function is chosen as displayed in Table 33.3. If you specify neither the `DIST=` option nor the `LINK=` option, then the CAUSALMED procedure defaults to the binary distribution with logit link if the outcome variable is listed in the `CLASS` statement. If the outcome variable is not listed in the `CLASS` statement, then the CAUSALMED procedure defaults to the normal distribution with the identity link function.

<table>
<thead>
<tr>
<th>DIST=</th>
<th>Distribution</th>
<th>Default Link Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIN</td>
<td>Binary</td>
<td>Logit</td>
</tr>
<tr>
<td>NEGBIN</td>
<td>Negative binomial</td>
<td>Log</td>
</tr>
<tr>
<td>NORMAL</td>
<td>Normal</td>
<td>Identity</td>
</tr>
<tr>
<td>POISSON</td>
<td>Poisson</td>
<td>Log</td>
</tr>
</tbody>
</table>

For the Poisson and negative binomial distributions, responses must be nonnegative, but they can take noninteger values. Observations whose response values are outside of the distribution’s support are not used to estimate the mediation effects.

**LINK=** *keyword*

specifies the link function in the model. You can specify the *keywords* shown in Table 33.4.

<table>
<thead>
<tr>
<th>LINK=</th>
<th>Link Function</th>
<th>g(μ) =</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDENTITY</td>
<td>Identity</td>
<td>μ</td>
</tr>
<tr>
<td>LOG</td>
<td>Log</td>
<td>log(μ)</td>
</tr>
<tr>
<td>LOGIT</td>
<td>Logit</td>
<td>log(μ/(1 − μ))</td>
</tr>
</tbody>
</table>

By default, the link function is chosen as shown in Table 33.3.
Causal Mediation Effects: Definitions, Assumptions, and Identification

This section describes the theoretical foundation of the CAUSALMED procedure. It defines the mediation effects and related effects that the procedure estimates, and it discusses the implications of the theoretical framework for valid application of the procedure.

In any causal mediation analysis, there are four main variables of interest:

- an outcome variable \( Y \)
- a treatment variable \( T \) that is hypothesized to have direct and indirect causal effects on the outcome variable \( Y \) (in epidemiology, a treatment variable is also known as an exposure, denoted as \( A \))
- a mediator variable \( M \) that is hypothesized to be causally affected by the treatment variable \( T \) and that itself has a direct effect on the outcome variable \( Y \)
- a set of pretreatment or background covariates that confound the observed relationships among \( Y, T, \) and \( M \)

Figure 33.8 represents the first three variables in a causal diagram. A causal diagram depicts the causal relationships of variables in an intuitive way. For a general theory of causal diagrams, see Pearl (2009). The role of the background covariates \( C \) is discussed after the causal diagram is interpreted.

![Figure 33.8 A Causal Mediation Model](image)

Figure 33.8 shows two causal pathways that represent the effect of \( T \) on \( Y \):

- A direct pathway: \( T \rightarrow Y \)
- A mediated or indirect pathway: \( T \rightarrow M \rightarrow Y \)

The first causal pathway generates the \textit{direct} effect of \( T \) on \( Y \), and the second pathway generates the \textit{indirect} effect of \( T \) on \( Y \).

Suppose that \( Y, T, \) and \( M \) are all continuous variables. If you ignore the causal pathways and regress \( Y \) on \( T \) by using a linear model of the form

\[
Y = \gamma_0 + \gamma_1 T + e
\]
where $e$ is an error term that has an expected value of 0 and $\gamma_0$ is an intercept, then $\gamma_1$ is referred to as the total effect of $T$ on $Y$. This total effect is the overall effect of $T$ on $Y$ without referring to a particular pathway.

When you hypothesize a causal diagram such as Figure 33.8, the relationships among $Y$, $T$, and $M$ are described by two linear equations,

\[
\begin{align*}
M &= \beta_0 + \beta_1 T + \epsilon \\
Y &= \theta_0 + \theta_1 T + \theta_2 M + \delta
\end{align*}
\]

where $\epsilon$ and $\delta$ are error terms that have expected values of 0, and the parameters of these two equations are as follows:

- $\beta_0$ is the intercept of the equation for predicting $M$.
- $\theta_0$ is the intercept of the equation for predicting $Y$.
- $\beta_1$ is the effect of the $T \rightarrow M$ path.
- $\theta_1$ is the effect of the $T \rightarrow Y$ path.
- $\theta_2$ is the effect of the $M \rightarrow Y$ path.

Substituting the equation for predicting $M$ into that for predicting $Y$, you have

\[Y = (\theta_0 + \theta_2 \beta_0) + (\theta_1 + \theta_2 \beta_1)T + (\theta_2 \epsilon + \delta)\]

Comparing this equation with the regression equation that predicts $Y$ by $T$ ignoring the causal pathways, you have the equality

\[\gamma_1 = \theta_1 + \beta_1 \theta_2\]

where the two terms on the right side of the equation represent additive components of the total effect $\gamma_1$, assuming that the causal diagram and the corresponding linear equations are true.

Because the first component $\theta_1$ represents the direct effect of the $T \rightarrow Y$ path, the second component $\theta_2 \beta_1$ thus represents the effect of $T$ on $Y$ that is not direct, or simply the indirect effect of $T$ on $Y$. You can also interpret this indirect effect $(\beta_1 \theta_2)$ intuitively—it is the product of the two path effects along the indirect pathway $T \rightarrow M \rightarrow Y$.

Therefore, conceptually, the total effect decomposition can be written as follows:

\[\text{total effect} = \text{direct effect} + \text{indirect effect}\]

The direct and indirect effect components are also well defined by the parameters in linear models for continuous $Y$, $T$, and $M$. For an illustration, see Example 33.4.

However, the illustration of the total effect decomposition has been quite ad hoc in nature. It is based on comparing linear models for continuous variables without prior definitions of direct and indirect effects. Consequently, for nonlinear models or linear models that have interaction effects between $T$ and $M$, the preceding strategy would not work. One reason is that there could be more than two terms in the decomposition so
that the direct-indirect decomposition is ambiguous. Another reason is that the terms become much more complicated in nonlinear models, and how to obtain those direct-indirect components would not be clear.

In contrast, the counterfactual framework addresses this issue by offering clear definitions of direct and indirect effects that are applicable to linear and nonlinear models with or without interaction effects. The next section describes this framework.

Another limitation of the illustration that is based solely on the diagram in Figure 33.8 is that it does not deal adequately with pretreatment characteristics or covariates C in observational studies. Typically, covariates C functions like common causes among Y, T, and M in a causal diagram. In observational studies, the observed associations or relationships among Y, T, and M are attributed to two parts. One part is the actual causal effects among them (that is, the effects that are due to the previously mentioned direct and indirect causal pathways). The other part is their induced associations by C. This part of induced association is often called confounding associations or effects. To obtain unbiased estimates of causal mediation and related effects in observation studies, statistical methods must be able to “remove” the confounding associations.

Before a discussion of such statistical methods, a more fundamental issue needs to be addressed: Under what conditions can causal mediation effects be identified? Only after the identification conditions are satisfied can you then attempt to obtain unbiased estimation of causal mediation and related effects. The identification issue is addressed in the section “Identification of Causal Mediation Effects” on page 2102 after the counterfactual framework is described in the next section. Regression adjustment methods that are based on the identification conditions are then presented in the section “Regression Methods for Causal Mediation Analysis” on page 2103.

**Counterfactual Framework for Defining Causal Mediation Effects**

Mediation analysis has a relatively long history in the field of psychology. Almost all recent developments in the area of causal mediation analysis trace back to the psychological tradition of mediation analysis, as typified by Baron and Kenny (1986). The preceding section illustrates such a traditional approach.

However, as discussed in the preceding section, a problem of the traditional approach is that it lacks a general framework that offers clear definitions of causal mediation and related effects. As a result, the traditional approach cannot deal with interaction effects effectively and it cannot treat binary outcomes and binary mediators in a unified framework.

The counterfactual framework (Robins and Greenland 1992; Pearl 2001) offers a solution to this problem. Within this framework, direct and indirect effects are well defined in terms of counterfactual outcomes. Using these definitions, VanderWeele and Vansteelandt (2009) and VanderWeele and Vansteelandt (2010) derived analytic results for computing causal mediation effects under a wide class of parametric models for various types of treatment and outcome variables. Valeri and VanderWeele (2013) extended these results to binary mediators and count outcomes. This line of development provides the theoretical foundation for the CAUSALMED procedure.

A counterfactual outcome is the outcome that you would observe under a hypothetical intervention that you can set the treatment T to particular level t. Counterfactual outcomes, which are also called potential outcomes by some researchers, are therefore defined for scenarios that might be contrary to the factual outcomes. In the counterfactual framework for causal mediation analysis, interventions on the mediator level are also used in various hypothetical scenarios for defining mediation effects.

The following notation is used for counterfactual outcomes that depend on interventions:
- $Y_t$ is the counterfactual outcome of $Y$ for a subject when an intervention sets the treatment level to $T = t$.
- $M_t$ is the counterfactual outcome of $M$ for a subject when an intervention sets the treatment level to $T = t$.
- $Y_{tm}$ is the counterfactual outcome of $Y$ for a subject when an intervention sets the treatment level to $T = t$ and $M = m$.

This notation places no restriction on variable types. The variables $Y$, $T$, and $M$ can be continuous or binary.

Suppose for the moment that the treatment is binary so that $t$ is either 0 or 1, denoting the control (no treatment) and treatment conditions, respectively. The total effect (TE) for a subject is defined as the difference between the counterfactual outcomes at the treatment and control levels:

$$\text{TE} = Y_{1M_1} - Y_{0M_0}$$

In this equation, the first subscript in the counterfactual outcomes denotes the intervention of the treatment (either at 1 or 0), and the second subscript denotes the mediator value that would follow from the intervention of the treatment (either $M_1$ or $M_0$).

The controlled direct effect (CDE) for a subject is defined as the difference between the counterfactual outcomes at the two treatment levels when an intervention sets the mediator to a particular level $M = m$. That is,

$$\text{CDE}(m) = Y_{1m} - Y_{0m}$$

The natural direct effect (NDE) for a subject is defined as the difference between the counterfactual outcomes at the two treatment levels when an intervention sets the mediator value to $M = M_0$, which is the natural level of the mediator when there is no treatment. That is,

$$\text{NDE} = Y_{1M_0} - Y_{0M_0}$$

The natural indirect effect (NIE) for a subject is defined as the difference between the counterfactual outcomes at the two mediator levels at $M_1$ and $M_0$ when an intervention sets the treatment to $T = 1$. That is,

$$\text{NIE} = Y_{1M_1} - Y_{1M_0}$$

All the preceding definitions assume that the treatment variable $T$ is binary. If the treatment variable is continuous, then the treatment levels must be defined according to the treatment and control levels of interest. For example, if $t_1$ and $t_0$ are the treatment and control levels on a continuous scale and they represent the levels of substantive interest, they should replace the 1 and 0 values, respectively, for the treatment and control levels in the definitions. However, this more general notation is not used here because it would make the presentation unnecessarily complicated.

These definitions have two important properties. First, they lead to the following conventional two-way decomposition of the total effect (TE):

$$\text{TE} = \text{NDE} + \text{NIE}$$
Second, these definitions are independent of the models for the outcome or mediator. Hence, these definitions and the total effect decomposition are applicable to linear or nonlinear models, with or without an interaction effect between $T$ and $M$.

The percentage of total effect that is mediated (PM) is computed as

$$\text{PM} = \frac{\text{NIE}}{\text{TE}} \times 100\%$$

VanderWeele (2014) took a step further and introduces the following four-way decomposition of the total effect:

$$\text{TE} = \text{CDE} + \text{IRF} + \text{IMD} + \text{PIE}$$

The component effects in this equation are called the controlled direct effect, the reference interaction, the mediated interaction (in VanderWeele (2014)), and the pure indirect effect, respectively. In VanderWeele (2014), IRF is denoted as $\text{INT}_{\text{ref}}$ and IMD is denoted as $\text{INT}_{\text{med}}$. These four component effects are also defined in terms of counterfactual outcomes. For definitions, see VanderWeele (2014).

The significance of these components in causal mediation analysis is that they characterize interaction and mediation effects as follows:

- CDE (controlled direct effect) is the component effect that is not due to interaction or mediation.
- IRF (reference interaction) is the component effect that is due to interaction but not mediation.
- IMD (mediated interaction) is the component effect that is due to both interaction and mediation.
- PIE (pure indirect effect) is the component effect that is due to mediation but not interaction.

Dividing each of these component effects by the total effect yields the corresponding proportion contributions of these components. However, these contributions are not interpretable when the components effects have mixed signs.

An important relationship between the two-way decomposition and the four-way decomposition is expressed by the following equation:

$$\text{NIE} = \text{PIE} + \text{IMD}$$

This equation expresses the mediation effect or natural indirect effect (NIE) as the composite component of the pure indirect effect and mediated interaction.

Another useful composite component of the four-way decomposition is the “portion attributed to interaction,” which is defined as

$$\text{PAI} = \text{IRF} + \text{IMD}$$

As its name suggests, this is the portion of the total effect that is due to the interaction between $T$ and $M$. The percentage of total effect that is due to the interaction is therefore computed as

$$\text{PAI/TE} \times 100\%$$

VanderWeele (2014) discusses various two-way and three-way decompositions and their relationships with the four-way decomposition. He also offers interesting interpretations and applications of these decompositions. For any causal mediation analysis, you can use the DECOMP option in the CAUSALMED procedure to obtain several two-way decompositions, several three-way decompositions, and the four-way decomposition. For more information about the decompositions, see the DECOMP option.
### Identification of Causal Mediation Effects

This section lays out the identification conditions of causal mediation effects and their implications for applying statistical methods that aim to obtain unbiased estimation of the effects.

First, it is useful to distinguish the following three types of confounding covariates:

- $C_1$ represents a generic covariate that confounds the relationship between $T$ and $Y$. This is a treatment-outcome confounder.
- $C_2$ represents a generic covariate that confounds the relationship between $M$ and $Y$. This is a mediator-outcome confounder.
- $C_3$ represents a generic covariate that confounds the relationship between $T$ and $M$. This is a treatment-mediator confounder.

As in preceding sections, let $C$ denote any of the covariates $C_1$, $C_2$, or $C_3$. Thus, controlling for $C$ in regression analysis means that all types of confounding covariates are being controlled for.

According to Valeri and VanderWeele (2013), the following four assumptions are required for the identification of causal mediation effects:

- no unmeasured treatment-outcome confounders given $C$
- no unmeasured mediator-outcome confounders given $(C, T)$
- no unmeasured treatment-mediator confounders given $C$
- no mediator-outcome confounder is affected by $T$ (directly or indirectly) given $C$

The identification of the controlled direct effect (CDE) assumes the first two conditions, and the identification of the natural direct effect (NDE) and the natural indirect effect (NIE) assumes all four conditions. These four assumptions are collectively called the “no unmeasured confounding assumption.” Formal statements for these identification conditions can be found in the appendix of Valeri and VanderWeele (2013) and VanderWeele (2015).

Essentially, in order to obtain unbiased estimation of causal mediation and related effects, the regression adjustment method that is discussed in the next section assumes that the identification conditions are satisfied.

In practice, the implication is that in order to have valid causal interpretations of the mediation effects, you must be able to measure all relevant confounding covariates $C$ and include them in a causal mediation analysis. For example, the first identification condition states that there are no unmeasured treatment-outcome confounders given $C$. Practically, a simple interpretation of this condition is that if there are treatment-outcome confounders $C_1$ in the observational study, your set of $C$ must have measured and included these confounders in the analysis in order to obtain unbiased estimation of causal effects. Similarly, other identification conditions require $C_2$ or $C_3$, if present, to be measured and included in the analysis.
Regression Methods for Causal Mediation Analysis

The CAUSALMED procedure implements regression methods for estimating causal mediation effects that assume the identification conditions of the preceding section along with correct specification of the following two models:

- the outcome model for $Y$ given $T$, $M$, and $C$
- the mediator model for $M$ given $T$ and $C$

For a class of generalized linear models, VanderWeele and Vansteelandt (2009), VanderWeele and Vansteelandt (2010), and Valeri and VanderWeele (2013) derived analytic formulas for computing various causal mediation effects for different variable types, including combinations of the following cases:

- outcome variable $Y$, which can be binary, continuous, or count
- treatment variable $T$, which can be binary or continuous
- mediator variable $M$, which can be binary or continuous
- covariates $C$, which can be categorical or continuous

PROC CAUSALMED implements these analytic formulas. For the case that has a binary outcome and a continuous mediator, the analytic formulas assume that the outcome $Y$ is a rare event (Valeri and VanderWeele 2013; VanderWeele 2014). If $Y$ is not rare, then the formulas are still valid if $Y$ is modeled by using a log link.

Let $\theta$ represent the vector that collects all parameters in the outcome and mediator models. Under the correct specification of regression models and the identification assumptions, the causal effects in a mediation analysis are functions of $\theta$ conditional on the covariate values. That is, a causal effect, which is denoted by $g_{ef}$, can be expressed as a function of $\theta$ given $C = c$,

$$g_{ef}(\theta | C = c),$$

where $c$ represents some fixed values for covariates $C$. For continuous outcomes, the mediation effects $g_{ef}(\theta | C = c)$ are defined on the original scale. For binary outcomes, the mediation effects $g_{ef}(\theta | C = c)$ are defined on the odds ratio or excess relative risk scale. For the formulas, see Valeri and VanderWeele (2013) and VanderWeele (2014).

Due to possible nonlinearity and inclusion of interaction terms in the model, a causal effect $g_{ef}(\theta | C = c)$, is different, in general, for different sets of covariate values. By default, PROC CAUSALMED computes $g_{ef}(\theta | C = c)$ with $c = c_0$ where $c_0$ is the sample mean value of $C$. This default setting provides “overall” measures of various causal mediation effects. It is consistent with the treatment of the SAS macros that are implemented by Valeri and VanderWeele (2013). For categorical covariates, this default computation still applies. The mean values of the categorical covariates are computed from the dummy-coded 0-1 values for categorical levels. Then these mean values are put into formulas for computing the overall causal mediation effects.

However, this does not mean that PROC CAUSALMED requires you to dummy-code the categorical covariates for analysis. The dummy coding and the averaging are done internally in the procedure.
Maximum Likelihood Estimation

For random samples, PROC CAUSALMED estimates causal mediation effects by the maximum likelihood method. The maximum likelihood estimate $\hat{\theta}$ of $\theta$ is first estimated for the outcome and mediator models. Then the maximum likelihood estimates of various causal mediation effects are simply computed as

$$g_{ef}(\hat{\theta} \mid C = c_0),$$

where $ef$ is the index for effects and $C = c_0$ is the average of the covariate values that are computed from the sample. For categorical covariates, this definition of $c_0$ assumes that the levels are dummy-coded as 0 and 1, which is done internally by PROC CAUSALMED.

Given the estimated covariance matrix for $\hat{\theta}$, the delta method is used to estimate the standard errors for the causal effects $g_{ef}(\hat{\theta} \mid C = c_0)$. In the computation of these estimates, the covariate values $c_0$ are treated as fixed values. For more information about the delta method for computing standard errors in this context, see VanderWeele and Vansteelandt (2009) or VanderWeele (2015).

Alternatively, you can use bootstrap techniques to compute standard error estimates and confidence intervals. For more information about the bootstrap method, see the BOOTSTRAP statement and the section “Bootstrap Methods” on page 2111.

As explained in the preceding sections, the evaluation of causal mediation effects depends on the levels of covariates. In addition to the overall causal mediation effects that are evaluated at $C = c_0$, you can provide particular covariate levels, say $C = c_1$, that are particularly meaningful to your research by using the EVALUATE statement. The maximum likelihood estimate is then $g_{ef}(\theta \mid C = c_1)$ and the standard error is computed similarly by the delta method. For more information about evaluating causal mediation effects, see the section “Evaluating Causal Mediation Effects” on page 2104. For an illustration, see Example 33.2.

Estimation of Various Total Effect Decompositions

Formulas for estimating the components of the four-way decomposition of the total effect (VanderWeele 2014) follow essentially the same logic that is described in the preceding sections. For continuous outcomes, the components of the four-way decomposition are computed on the original scales. For binary outcomes, the components of the four-way decomposition are computed on the odds ratio scale and the excess relative risk scale. These formulas are quite involved and are not presented here. For more information, see VanderWeele (2014).

In addition to the four-way decomposition, PROC CAUSALMED estimates the component effects of several other two-way and three-way decompositions by using the same analytic technique as that of the four-way decomposition. You can use the DECOMP= option in the EVALUATE or PROC CAUSALMED statement to request these decompositions.

To compute standard error estimates for these component effects and their percentage contribution, PROC CAUSALMED uses the delta method with analytic derivatives. Bootstrap methods are also available for computing standard errors and confidence intervals. For more information about bootstrap estimation, see the BOOTSTRAP statement and the section “Bootstrap Methods” on page 2111.

Evaluating Causal Mediation Effects

In general, the CAUSALMED procedure computes causal mediation effects and decompositions that are conditioned on specific levels of covariates. In addition, some of the causal mediation effects are defined
at specific levels (numerical or categorical) of treatment, control, and mediator variables. Therefore, it is important to understand how to set these variable levels for evaluating causal mediation effects that meet your research goals.

This section explains the roles of treatment, control, mediator, and covariate levels in defining and computing causal mediation effects. It shows how you can use the options in the \texttt{EVALUATE} statement to specify these levels, and it describes the default levels that the \texttt{CAUSALMED} procedure uses.

Suppose that $T$ represents a treatment variable that has a causal effect on an outcome variable $Y$. Furthermore, suppose that $M$ represents a mediator variable, which is affected by $T$ and has a causal effect on $Y$, and that $C$ represents a generic covariate that confounds the causal treatment and mediation effects.

The roles of the treatment, control, mediator, and covariate levels in defining causal mediation effects are as follows:

- The level $t_1$ of the treatment variable $T$ is the level that you designate as the treatment condition for all the effects and decompositions that are computed. For example, if the variable $T$ represents the dosage level of a drug, $t_1 = 10$ mg is the dosage level that defines the treatment condition. For a binary treatment variable, researchers usually define $t_1$ as 1 to represent the presence of the treatment.

- The level $t_0$ of the treatment variable $T$ is the level that you designate as the reference or control condition for all the effects and decompositions that are computed. For example, if the variable $T$ represents the dosage level of a drug, $t_0 = 5$ mg is the dosage level that defines the control condition. For a binary treatment variable, researchers usually define $t_0$ as 0 to represent the absence of treatment.

- The level $m^*$ of the mediator variable $M$ is the level that you designate to compute the controlled direct effect. For binary mediator variables, researchers usually define $m^*$ as 0 so that they can evaluate the controlled direct effect by holding the mediator value at the “absence” level.

- The levels $c$ of the covariates are the conditional covariate values in the formulas for computing causal mediation effects.

In general, specifying covariate levels $c$, the treatment level $t_1$ (of the treatment variable), or the control level $t_0$ (of the treatment variable) changes the estimates of all mediation effects and decompositions. Specifying the controlled level $m^*$ of the mediator variable does not change the estimates of the total effect (TE), the natural direct effect (NDE), or the natural indirect effect (NIE). But it does change the estimates of the controlled direct effect (CDE) and the reference interaction (IRF).

### Default Settings of Treatment and Control Levels

For binary treatment variables, \texttt{PROC CAUSALMED} uses the first level of the variable as the default treatment level and the second (last) level of the variable as the default control level. In other words, the first level of a binary treatment variable takes the role of $t_1$ and the second level takes the role of $t_0$.

For continuous or ordinal treatment variables, researchers habitually set levels of $t_1$ and $t_0$ in such a way that their difference is 1. Such a habitual setting serves well for linear models, including linear regression analysis and linear structural equation modeling. The associated regression coefficient (or effect) is defined as the change in the outcome $Y$ for a unit change in the predictor $T$. In linear models, the effect on $Y$ depends only on the \textit{difference} between $t_1$ and $t_0$ but not on the levels of $t_1$ and $t_0$ themselves.

However, with nonlinear models, binary responses, and interaction effects, the computation of causal mediation effects and decompositions does, in general, depend on the levels of $t_1$ and $t_0$. Using different
sets of $t_1$ and $t_0$ (even if their difference remains constant) leads to numerically different estimates of causal mediation effects. By default, PROC CAUSALMED sets the treatment and control levels around the center of the distribution of the treatment variable. That is,

$$t_1 = \bar{t} + 0.5$$
$$t_0 = \bar{t} - 0.5$$

where $\bar{t}$ is the sample mean of the treatment variable. This sample mean value is treated as fixed when computing standard errors.

You can define your own treatment and control levels for evaluating causal mediation effects and decompositions. For example, instead of using a single unstandardized unit as the treatment amount, you can use one standard deviation,

$$t_1 = \bar{t} + 0.5 \times s_t$$
$$t_0 = \bar{t} - 0.5 \times s_t$$

where $s_t$ is the sample standard deviation of the treatment variable. This sample standard deviation is treated as fixed when computing standard errors.

PROC CAUSALMED enables you to set the treatment and control levels either on an unstandardized scale or on a standardized scale. Table 33.5 presents more options for setting these levels.

### Default Settings of the Controlled Mediator Level

For binary mediator variables, PROC CAUSALMED uses the second (last) level of the variable as the default controlled (baseline) level, $m^*$, of the mediator variable. This is consistent with the way that you specify the mediator model in the MEDIATOR statement. That is, by default, the procedure models the probability of the event indicated by the first level of the mediator variable.

For continuous or ordinal mediator variables, PROC CAUSALMED uses the sample mean of $M$ as the default controlled mediator level, $m^*$, when evaluating causal mediation effects. Table 33.5 presents more options for setting this level.

### Covariate Levels and Their Default Settings

When you specify the effects of confounder covariates in the COVAR statement, the CAUSALMED procedure computes mediation effects conditionally at specific levels of the covariates. You can provide one or more EVALUATE statements to request that these effects be computed at specified settings that are of interest in your study. However, whether or not you provide an EVALUATE statement, the CAUSALMED procedure uses the sample means of the covariates to compute “overall” measures of causal mediation effects, which are displayed in the “Summary of Effects” table. For an illustration, see Example 33.2.

Although the means of ordinal and continuous covariates are well defined, less apparent is how to define the mean levels of categorical covariates and any interaction terms that might be included in the model.

To illustrate this, suppose that $C_1$ is a continuous covariate and $C_2$ is a categorical covariate that has three levels: 1, 2, and 3. Also suppose that there are six observations for $C_1$ and $C_2$:
All other variables are not shown.

The following design matrix for the linear predictor contains one column for C1, three columns for C2, and three columns for the interaction of the two variables:

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>C1 x C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1 0 0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2 0 0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0 3 0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0 4 0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0 5</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0 6</td>
</tr>
</tbody>
</table>

The parameterization shown here for C2 is represented internally in PROC CAUSALMED. You are not required to use this coding in the input.

The marginal means of the seven columns are 3.5, 1/3, 1/3, 1/3, 1.5, 3.5, and 5.5, respectively. By default, PROC CAUSALMED substitutes these means for covariate levels in the formulas for computing mediation effects and decompositions.

Substitution of marginal means makes intuitive sense when the models for Y and M are both linear. In this case, the computed causal mediated effects and decompositions can be interpreted as marginal effects. However, causal mediation effects that are computed in this way for nonlinear models (for example, binary responses with logit links) cannot be interpreted as marginal effects. Nonetheless, the default provides “overall” causal mediation effect estimates that are not entirely arbitrary. In a sense, the default method for categorical covariates provides an averaged categorical profile for evaluating causal mediation effects.

The default levels are not the only setting that you can consider. In this example, it would be interesting to conduct three causal mediation analyses, each of which is conditioned on a particular level of C2. You can request these analyses by specifying the following EVALUATE statements:

```plaintext
evaluate 'Conditional on Level 1 of C2' C1=mean C2='1';
evaluate 'Conditional on Level 2 of C2' C1=mean C2='2';
evaluate 'Conditional on Level 3 of C2' C1=mean C3='3';
```

Each EVALUATE statement generates a set of mediation analysis results.

In summary, you can use the EVALUATE statement to examine causal mediation effects that are conditional on the covariate levels that you specify. The CAUSALMED procedure displays these effects in the output together with overall effects that are conditioned on default settings; For illustrations, see Example 33.2 and Example 33.3. The next section describes the options for specifying treatment, control, mediator, and covariate levels.
Options for Setting Variables Levels

You use the `EVALUATE` statement to request the computation of causal mediation effects that are conditional on particular levels of variables. You can set the levels of variables by specifying an assignment of the following form:

```
var-key=value-key
```

Table 33.5 summarizes the options for `var-key` and `value-key`. The last two columns of Table 33.5 display the default `value-key`.

<table>
<thead>
<tr>
<th>Level</th>
<th>var-key</th>
<th>value-key</th>
<th>Default value-key</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class Variable</td>
<td>Count or Continuous Variable</td>
<td>Class Variable</td>
</tr>
<tr>
<td>Treatment</td>
<td>_TREATMENT FIRST</td>
<td>MAX</td>
<td>FIRST</td>
</tr>
<tr>
<td></td>
<td>_A1 LAST</td>
<td>MEAN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>_T1 'level'</td>
<td>MIN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vname(TREATMENT)</td>
<td>value</td>
<td>value(SD)</td>
</tr>
<tr>
<td>Control</td>
<td>_CONTROL FIRST</td>
<td>MAX</td>
<td>LAST</td>
</tr>
<tr>
<td></td>
<td>_A0 LAST</td>
<td>MEAN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>_T0 'level'</td>
<td>MIN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vname(CONTROL)</td>
<td>value</td>
<td>value(SD)</td>
</tr>
<tr>
<td>Mediator</td>
<td>_MEDIATOR FIRST</td>
<td>MAX</td>
<td>LAST</td>
</tr>
<tr>
<td></td>
<td>_MSTAR LAST</td>
<td>MEAN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vname 'level'</td>
<td>MIN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>value</td>
<td>value</td>
<td>value(SD)</td>
</tr>
<tr>
<td>Covariate</td>
<td>vname FIRST</td>
<td>MAX</td>
<td>mean</td>
</tr>
<tr>
<td></td>
<td>LAST</td>
<td>MEAN</td>
<td>or</td>
</tr>
<tr>
<td></td>
<td>MODE</td>
<td>MIN</td>
<td>MODE</td>
</tr>
<tr>
<td></td>
<td>'level'</td>
<td>value</td>
<td>value(SD)</td>
</tr>
</tbody>
</table>

In this table, `vname` represents an actual variable name, `'level'` represents an actual level of a classification variable, and `value` represents an actual value of a numeric variable. In the last two columns, `mean` represents the sample mean of a continuous variable or the sample mean of a categorical variable (in dummy coding).
To specify an assignment, first look for the correct var-key in the second column. Different var-keys are used for the treatment, control, mediator, and covariate levels. In all cases, you can use the actual variable name of the variable. Next, select one of the value-keys in the third or fourth column to specify the desired variable level.

Repeat as many assignments as you need to specify the levels of various variables.

For example, suppose that there is a continuous treatment variable Exposure and a binary mediator variable PerceivedPain in your analysis. You identify the roles of these variables by using the following statements:

```plaintext
proc causalmed;
  class PerceivedPain;
  mediator PerceivedPain = Exposure;
  model outcome = PerceivedPain | Exposure;
```

To set the treatment level at the maximum sample value, the control level at the mean value, and the mediator at the level encoded as “none,” you can use any of the following equivalent specifications:

```plaintext
evaluate 'Setting 1' _t1=max _t0=mean _mstar='none';
evaluate 'Setting 2' _treatment=max _control=mean _mediator='none';
evaluate 'Setting 3' Exposure(treatment)=max Exposure(control)=mean
  PerceivedPain='none';
```

This example shows that you can specify a var-key either directly (by providing an actual variable name) or indirectly (by providing a keyword). Likewise, you can specify a value-key either directly (by providing an actual level) or indirectly (by providing a keyword). For a complete description of these options, see the EVALUATE statement.

Note that the default value-key for categorical covariates can be either the sample means (denoted as mean in the table) or MODE. If you do not assign any levels for categorical covariates in an EVALUATE statement, PROC CAUSALMED uses the sample means as the default levels for all unassigned categorical covariates that are specified in the COVAR statement. For example, the sample means of C1, C2, and C3 are the default levels used in the EVALUATE statement for the following specification:

```plaintext
proc causalmed;
  class C1 C2 C3;
  mediator M = T;
  model Y = T | M;
  covar C1 C2 C3 C4;
  evaluate 'Conditional on C4=max' C4=max M=mean;
```

If you assign the level of at least one categorical covariate in an EVALUATE statement, PROC CAUSALMED uses MODE as the default level for the unassigned categorical covariates that are specified in the COVAR statement. For example, the modal levels of C2 and C3 and the sample mean of C4 are the default levels used in the EVALUATE statement for the following specification:

```plaintext
proc causalmed;
  class C1 C2 C3;
  mediator M = T;
  model Y = T | M;
  covar C1 C2 C3 C4;
  evaluate 'Conditional on C1=1' C1='1' M=mean;
```
Multimodal Covariates

If you specify MODE as the value-key for a categorical covariate and it has multiple modes, an averaging process is used to compute the levels. To illustrate this, suppose that C1 is a continuous covariate and C2 and C3 are binary covariates. Also suppose that there are six observations with the following values for the three covariates:

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The design matrix for the linear predictor contains one column for C1 and two columns for each of C2 and C3:

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Suppose you specify the following EVALUATE statement:

```
evaluate 'Setting A' C1=mean C2=mode C3=mode;
```

The mean of C1 is 3.5. The modal class of C2 is ‘1’, and hence the coding ‘1 0’ is used as the covariate level for C2. However, because C3 has two modal classes, ‘1 0’ and ‘0 1’, these two modal class codings are averaged out with other levels. The final coding vector for the covariate levels is then the average of the following two vectors:

| 3.5 | 1  | 0  | 0  |
| 3.5 | 1  | 0  | 1  |

As a result, the averaged levels 3.5, 1, 0, 0.5, and 0.5 are used in the formulas for evaluating causal mediation effects and decompositions.

If an interaction between C1 and C3 is also modeled, then the average of the following two vectors is used:

| 3.5 | 1  | 0  | 1  | 0  | 3.5 | 0  |
| 3.5 | 1  | 0  | 0  | 1  | 0  | 3.5|

Here the last two columns represent the interaction terms. As a result, the averaged levels 3.5, 1, 0, 0.5, 0.5, 1.75, and 1.75 are used in the formulas for evaluating causal mediation effects and decompositions.
Bootstrap Methods

If you specify the `BOOTSTRAP` statement, PROC CAUSALMED uses bootstrap resampling to compute standard errors and confidence intervals for causal mediation effects and decompositions. The procedure samples as many bootstrap sample data sets (replicates) as you specify in the `NBOOT=` option and then estimates the effects and decompositions for each replication.

Bootstrap confidence intervals are computed only for the effects and their corresponding percentages. These intervals are not computed for the parameters in the outcome or mediator models. You can specify one or more of the following types of bootstrap confidence intervals by using the `BOOTCI` option in the `BOOTSTRAP` statement:

- The `BOOTCI(NORMAL)` option requests bootstrap confidence intervals that are based on the normal approximation method. The $(1 - \alpha)100\%$ normal bootstrap confidence interval is given by
  \[
  \hat{\mu}_j \pm \sigma_{\mu_j} \times z_{(1 - \alpha/2)}
  \]
  where \( \hat{\mu}_j \) is the estimate of \( \mu_j \) from the original sample, \( \sigma_{\mu_j} \) is the standard deviation of the bootstrap parameter estimates, and \( z_{(1 - \alpha/2)} \) is the 100\%$(1 - \alpha/2)$th percentile of the standard normal distribution.

- The `BOOTCI(PERC)` option requests bootstrap confidence intervals that are based on the percentile method. The confidence limits are the 100\%$\alpha/2$th and 100\%$(1 - \alpha/2)$th percentiles of the bootstrap parameter estimates, which are computed as follows. Let \( \mu_{j,1}^*, \mu_{j,2}^*, \ldots, \mu_{j,B}^* \) represent the ordered values of the bootstrap estimates for the potential outcome mean \( \mu_j \). Let the \( k \)th weighted average percentile be \( q \), set \( p = \frac{k}{100} \), and let
  \[
  np = l + g
  \]
  where \( l \) is the integer part of \( np \) and \( g \) is the fractional part of \( np \). Then the \( k \)th percentile, \( q \), is computed as follows, which corresponds to the default percentile definition used by the UNIVARIATE procedure:

  \[
  q = \begin{cases} 
  \frac{1}{2}(\mu_{j,l}^* + \mu_{j,l+1}^*) & \text{if } g = 0 \\
  \mu_{j,l+1}^* & \text{if } g > 0 
  \end{cases}
  \]

- The `BOOTCI(BC)` option requests bias-corrected bootstrap confidence intervals, which use the cumulative distribution function (CDF), \( G(\mu^*) \), of the bootstrap parameter estimates to determine the upper and lower endpoints of the confidence interval. The bias-corrected bootstrap confidence interval is given by
  \[
  G^{-1}(\Phi(2z_0 \pm z_{\alpha/2}))
  \]
  where \( \Phi \) is the standard normal CDF, \( z_{\alpha/2} = \Phi^{-1}(\alpha/2) \), and \( z_0 \) is a bias correction,

  \[
  z_0 = \Phi^{-1}(\frac{N(\mu_j^* \leq \hat{\mu}_j)}{B})
  \]
  where \( \hat{\mu}_j \) is the original sample estimate of \( \mu_j \) from the input data set, \( N(\mu_j^* \leq \hat{\mu}_j) \) is the number of bootstrap estimates \( \mu_j^* \) that are less than or equal to \( \hat{\mu}_j \), and \( B \) is the number of bootstrap replicates for which an estimate for the treatment effect is obtained.
Bias-corrected bootstrap confidence intervals are the default.

PROC CAUSALMED requires at least 50 bootstrap samples for normal bootstrap confidence intervals and does not compute them if fewer than 40 of the samples produce usable estimates. The procedure requires at least 1,000 bootstrap samples for percentile and bias-corrected bootstrap confidence intervals and does not compute them if fewer than 900 of the samples produce usable estimates. If the number of samples $n$ specified in the NBOOT=$n$ option is less than 1,000 and percentile or bias-corrected bootstrap confidence intervals are requested, the value of $n$ is ignored.

### ODS Table Names

PROC CAUSALMED assigns a name to each table it creates. You can use these names to refer to the table when you use the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in Table 33.6. The options or specifications of the specific statements that produce these output tables are shown in the last two columns. For more information about ODS, see Chapter 20, “Using the Output Delivery System.”

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Statement</th>
<th>Option or Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classlevels</td>
<td>Classification variable levels</td>
<td>CLASS</td>
<td>Class variables</td>
</tr>
<tr>
<td>EffectDecomp</td>
<td>Decompositions of the total effect</td>
<td>PROC</td>
<td>DECOMP</td>
</tr>
<tr>
<td>EffectSummary</td>
<td>Summary of the direct and mediated effects</td>
<td></td>
<td>Default</td>
</tr>
<tr>
<td>MediatorProfile</td>
<td>Frequency counts for a binary mediator variable</td>
<td>CLASS</td>
<td>Binary mediator</td>
</tr>
<tr>
<td>MediatorEstimates</td>
<td>Parameter estimates for the mediator model</td>
<td>PROC</td>
<td>PMEDMOD</td>
</tr>
<tr>
<td>ModelInfo</td>
<td>Model information</td>
<td></td>
<td>Default</td>
</tr>
<tr>
<td>NObs</td>
<td>Number of observations</td>
<td></td>
<td>Default</td>
</tr>
<tr>
<td>OutcomeEstimates</td>
<td>Parameter estimates for the outcome model</td>
<td>PROC</td>
<td>POUTCOMEMOD</td>
</tr>
<tr>
<td>PercentDecomp</td>
<td>Percentage decompositions of the total effect</td>
<td>PROC</td>
<td>DECOMP</td>
</tr>
<tr>
<td>ResponseProfile</td>
<td>Frequency counts for a binary outcome variable</td>
<td>MODEL</td>
<td>DIST=BIN</td>
</tr>
<tr>
<td>TreatmentProfile</td>
<td>Frequency counts for a binary treatment variable</td>
<td>CLASS</td>
<td>Binary treatment</td>
</tr>
</tbody>
</table>

To control the display of multiple ODS tables, you can override the “Default” settings in Table 33.6 by specifying global display options in the PROC CAUSALMED statement. Table 33.7 shows these options. The ODS tables that are displayed by these options are marked by *. Notice that the NOPRINT option suppresses all ODS table output.
### Table 33.7  Global Display Options of the CAUSALMED Procedure

<table>
<thead>
<tr>
<th>Options</th>
<th>PALL</th>
<th>Default</th>
<th>PSHORT</th>
<th>PSUMMARY</th>
<th>NOPRINT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classlevels</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EffectDecomp</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EffectSummary</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>MediatorProfile</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MediatorEstimates</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ModelInfo</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NObs</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OutcomeEstimates</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PercentDecomp</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResponseProfile</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TreatmentProfile</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Examples: CAUSALMED Procedure

#### Example 33.1: Mediation Analysis with Interaction Effects and Four-Way Decomposition

This example continues the example in the section “Getting Started: CAUSALMED Procedure” on page 2075 by including an interaction effect between the treatment and the mediator in the outcome model. It also shows how you can obtain a four-way decomposition of the effects.

The goals of the observational study on page 2075 are to determine whether an encouraging environment provided by parents (which is represented by the variable Encourage) has an effect on the cognitive development of children (which is represented by the variable CogPerform) and to estimate the amount of the total causal effect that is due to the mediation of learning motivation (which is represented by the variable Motivation).

In the example on page 2075, the following statements are used to request a mediation analysis in which the main effects of Encourage and Motivation are specified in the outcome model:

```sas
proc causalmed data=Cognitive;
   model CogPerform = Encourage Motivation;
   mediator Motivation = Encourage;
   covar FamSize SocStatus;
run;
```

This analysis also specifies confounding covariates, and it produces the summary of effects in Output 33.1.1.
Output 33.1.1  Estimation of Causal Effects Adjusting for Confounding Covariates

<table>
<thead>
<tr>
<th>Summary of Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z  Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect</td>
<td>6.8421</td>
<td>0.1430</td>
<td>6.5618</td>
<td>7.1224</td>
<td>47.84</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876</td>
<td>4.2717</td>
<td>89.00</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
<td>4.1509</td>
<td>0.0471</td>
<td>4.0587</td>
<td>4.2432</td>
<td>88.21</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
<td>2.6912</td>
<td>0.1453</td>
<td>2.4065</td>
<td>2.9759</td>
<td>18.53</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>39.3325</td>
<td>1.3704</td>
<td>36.6465</td>
<td>42.0184</td>
<td>28.70</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>0.4197</td>
<td>0.0237</td>
<td>0.3733</td>
<td>0.4661</td>
<td>17.73</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>38.9128</td>
<td>1.3574</td>
<td>36.2524</td>
<td>41.5733</td>
<td>28.67</td>
</tr>
</tbody>
</table>

The following statements extend the analysis by including an interaction term between Encourage and Motivation in the outcome model:

```plaintext
proc causalmed data=Cognitive decomp;
  model CogPerform = Encourage | Motivation;
  mediator Motivation = Encourage;
  covar FamSize SocStatus;
run;
```

The specification `Encourage | Motivation` includes the main effects of Encourage and Motivation and their interaction. Equivalently, you could specify `Encourage Motivation Encourage*Motivation`, where the third term represents the interaction effect. The results of this analysis are shown in Output 33.1.2.

Output 33.1.2  Summary of Causal Effects with Interaction Effects

<table>
<thead>
<tr>
<th>Summary of Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z  Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect</td>
<td>6.8421</td>
<td>0.1430</td>
<td>6.5618</td>
<td>7.1224</td>
<td>47.84</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876</td>
<td>4.2717</td>
<td>89.00</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
<td>4.1509</td>
<td>0.0471</td>
<td>4.0587</td>
<td>4.2432</td>
<td>88.21</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
<td>2.6912</td>
<td>0.1453</td>
<td>2.4065</td>
<td>2.9759</td>
<td>18.53</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>39.3325</td>
<td>1.3704</td>
<td>36.6465</td>
<td>42.0184</td>
<td>28.70</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>0.4197</td>
<td>0.0237</td>
<td>0.3733</td>
<td>0.4661</td>
<td>17.73</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>38.9128</td>
<td>1.3574</td>
<td>36.2524</td>
<td>41.5733</td>
<td>28.67</td>
</tr>
</tbody>
</table>

When the interaction term is included, the ‘Percentage Mediated’ changes slightly from 37% (for the model without this term) to 39%. Although the percentage due to interaction that is shown in Output 33.1.2 is significant, it is less than 1%. Therefore, the interpretation of the results is not drastically different from those of the analysis with no interaction.
When you specify the DECOMP option, the CAUSALMED procedure generates a table, shown in Output 33.1.3, that displays various decompositions of the total effect: several two-way and three-way decompositions and a four-way decomposition. For more information about various decompositions and their interpretations, see the section “Causal Mediation Effects: Definitions, Assumptions, and Identification” on page 2097.

**Output 33.1.3** Decompositions of the Total Effect

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDE+NIE</td>
<td>Natural Direct</td>
<td>4.1509</td>
<td>0.0471</td>
<td>4.0587</td>
<td>4.2432</td>
<td>88.21</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Natural Indirect</td>
<td>2.6912</td>
<td>0.1453</td>
<td>2.4065</td>
<td>2.9759</td>
<td>18.53</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>CDE+PE</td>
<td>Controlled Direct</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876</td>
<td>4.2717</td>
<td>89.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Portion Eliminated</td>
<td>2.6625</td>
<td>0.1438</td>
<td>2.3807</td>
<td>2.9443</td>
<td>18.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>TDE+PIE</td>
<td>Total Direct</td>
<td>4.2084</td>
<td>0.0470</td>
<td>4.1163</td>
<td>4.3004</td>
<td>89.63</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>2.6338</td>
<td>0.1423</td>
<td>2.3548</td>
<td>2.9127</td>
<td>18.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>NDE+PIE+IMD</td>
<td>Natural Direct</td>
<td>4.1509</td>
<td>0.0471</td>
<td>4.0587</td>
<td>4.2432</td>
<td>88.21</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>2.6338</td>
<td>0.1423</td>
<td>2.3548</td>
<td>2.9127</td>
<td>18.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Mediated Interaction</td>
<td>0.0574</td>
<td>0.00339</td>
<td>0.05078</td>
<td>0.06407</td>
<td>16.93</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>CDE+PIE+PAI</td>
<td>Controlled Direct</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876</td>
<td>4.2717</td>
<td>89.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>2.6338</td>
<td>0.1423</td>
<td>2.3548</td>
<td>2.9127</td>
<td>18.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Portion Due to Interaction</td>
<td>0.0287</td>
<td>0.00201</td>
<td>0.02478</td>
<td>0.03265</td>
<td>14.30</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Four-Way</td>
<td>Controlled Direct</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876</td>
<td>4.2717</td>
<td>89.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Reference Interaction</td>
<td>-0.0287</td>
<td>0.00201</td>
<td>-0.03265</td>
<td>-0.02478</td>
<td>-14.30</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Mediated Interaction</td>
<td>0.0574</td>
<td>0.00339</td>
<td>0.05078</td>
<td>0.06407</td>
<td>16.93</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>2.6338</td>
<td>0.1423</td>
<td>2.3548</td>
<td>2.9127</td>
<td>18.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Total</td>
<td>Total Effect</td>
<td>6.8421</td>
<td>0.1430</td>
<td>6.5618</td>
<td>7.1224</td>
<td>47.84</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Note: NDE=CDE+IRF, NIE=PIE+IMD, PAI=RF+IMD, PE=PAI+PIE, TDE=CDE+PAI.

Important effects such as CDE, NDE, and NIE that contribute to the components of the decompositions are shown in the “Summary of Effects” table in Output 33.1.2.

The primary decomposition in the table in Output 33.1.3 is the four-way decomposition. All other component effects can be deduced by summing up particular subsets of the effects in the four-way decomposition. The note at the bottom of the table shows how component effects are related.

The DECOMP option also generates a table of percentage decompositions, which is shown in Output 33.1.4.
Chapter 33: The CAUSALMED Procedure

Output 33.1.4  Percentage Decompositions of the Total Effect

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>Effect</th>
<th>Percent</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z  Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Natural Direct</td>
<td>60.67</td>
<td>1.37</td>
<td>57.98</td>
<td>63.35</td>
<td>44.27 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Natural Indirect</td>
<td>39.33</td>
<td>1.37</td>
<td>36.65</td>
<td>42.02</td>
<td>28.70 &lt;.0001</td>
</tr>
<tr>
<td>CDE+PE</td>
<td>Controlled Direct</td>
<td>61.09</td>
<td>1.36</td>
<td>58.43</td>
<td>63.75</td>
<td>45.00 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Portion Eliminated</td>
<td>38.91</td>
<td>1.36</td>
<td>36.25</td>
<td>41.57</td>
<td>28.67 &lt;.0001</td>
</tr>
<tr>
<td>TDE+PIE</td>
<td>Total Direct</td>
<td>61.51</td>
<td>1.34</td>
<td>58.87</td>
<td>64.14</td>
<td>45.75 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>38.49</td>
<td>1.34</td>
<td>35.86</td>
<td>41.13</td>
<td>28.63 &lt;.0001</td>
</tr>
<tr>
<td>NDE+PIE+IMD</td>
<td>Natural Direct</td>
<td>60.67</td>
<td>1.37</td>
<td>57.98</td>
<td>63.35</td>
<td>44.27 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>38.49</td>
<td>1.34</td>
<td>35.86</td>
<td>41.13</td>
<td>28.63 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Mediated Interaction</td>
<td>0.84</td>
<td>0.04</td>
<td>0.77</td>
<td>0.91</td>
<td>23.65 &lt;.0001</td>
</tr>
<tr>
<td>CDE+PIE+PAI</td>
<td>Controlled Direct</td>
<td>61.09</td>
<td>1.36</td>
<td>58.43</td>
<td>63.75</td>
<td>45.00 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>38.49</td>
<td>1.34</td>
<td>35.86</td>
<td>41.13</td>
<td>28.63 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Portion Due to Interaction</td>
<td>0.42</td>
<td>0.02</td>
<td>0.37</td>
<td>0.47</td>
<td>17.73 &lt;.0001</td>
</tr>
<tr>
<td>Four-Way</td>
<td>Controlled Direct</td>
<td>61.09</td>
<td>1.36</td>
<td>58.43</td>
<td>63.75</td>
<td>45.00 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Reference Interaction</td>
<td>-0.42</td>
<td>0.02</td>
<td>-0.47</td>
<td>-0.37</td>
<td>-17.66 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Mediated Interaction</td>
<td>0.84</td>
<td>0.04</td>
<td>0.77</td>
<td>0.91</td>
<td>23.65 &lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>38.49</td>
<td>1.34</td>
<td>35.86</td>
<td>41.13</td>
<td>28.63 &lt;.0001</td>
</tr>
</tbody>
</table>

Note: NDE=CDE+IRF, NIE=PIE+IMD, PAI=IRF+IMD, PE=PAI+PIE, TDE=CDE+PAI.

This table shows that the two components that involve the interaction make up a very small percentage of the total effect. The mediated interaction effect represents less than 1%, and the reference interaction is very small and negative.

Example 33.2: Evaluating Controlled Direct Effects and Conditional Mediation Effects

This example continues the analysis of Example 33.1; it illustrates the use of the EVALUATE statement for computing controlled direct effects and mediation effects conditional on covariate values.

The following code includes three EVALUATE statements that assign different values for the mediator Motivation:

```plaintext
proc causalmed data=Cognitive;
  model CogPerform = Encourage | Motivation;
  mediator Motivation = Encourage;
  covar FamSize SocStatus;
  evaluate 'Default Mean Value of Mediator' Motivation=mean;
  evaluate 'High-Motivation Group' Motivation = 1(SD);
  evaluate 'Low-Motivation Group' Motivation = -1(SD);
run;
```

The labels, which are enclosed in quotation marks, distinguish the three EVALUATE statements and the output that they produce. The first EVALUATE statement specifies the level of the mediator Motivation as its mean. This happens to be the default level, so you should expect this statement to produce the same evaluation of causal effects that PROC CAUSALMED produces by default. Output 33.2.1 displays the causal
mediation effects that are evaluated by default, and Output 33.2.2 displays the causal mediation effects that are evaluated for the mediator level in the first EVALUATE statement. Clearly, these two tables are identical.

**Output 33.2.1** Summary of Effects (Default)

<table>
<thead>
<tr>
<th>Summary of Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z  Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect</td>
<td>6.8421</td>
<td>0.1430</td>
<td>6.5618 7.1224 47.84 47.84 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876 4.2717 89.00 89.00 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
<td>4.1509</td>
<td>0.0471</td>
<td>4.0587 4.2432 88.21 88.21 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
<td>2.6912</td>
<td>0.1453</td>
<td>2.4065 2.9759 18.53 18.53 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>39.3325</td>
<td>1.3704</td>
<td>36.6465 42.0184 28.70 28.70 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>0.4197</td>
<td>0.0237</td>
<td>0.3733 0.4661 17.73 17.73 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>38.9128</td>
<td>1.3574</td>
<td>36.2524 41.5733 28.67 28.67 &lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

**Output 33.2.2** Replicating the Default Summary of Effects

<table>
<thead>
<tr>
<th>Summary of Effects: Default Mean Value of Mediator</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z  Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect</td>
<td>6.8421</td>
<td>0.1430</td>
<td>6.5618 7.1224 47.84 47.84 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876 4.2717 89.00 89.00 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
<td>4.1509</td>
<td>0.0471</td>
<td>4.0587 4.2432 88.21 88.21 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
<td>2.6912</td>
<td>0.1453</td>
<td>2.4065 2.9759 18.53 18.53 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>39.3325</td>
<td>1.3704</td>
<td>36.6465 42.0184 28.70 28.70 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>0.4197</td>
<td>0.0237</td>
<td>0.3733 0.4661 17.73 17.73 &lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>38.9128</td>
<td>1.3574</td>
<td>36.2524 41.5733 28.67 28.67 &lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

There might be situations in which you want to evaluate the causal effects at other mediator levels. The second and third EVALUATE statements set the mediator level at one standard deviation above and below the mean, respectively. PROC CAUSALMED computes the sample mean and standard deviation (SD) for Motivation and then computes the levels of motivation as

\[ m^* = \text{mean} + \text{SD} \]

\[ m^* = \text{mean} - \text{SD} \]

These values of \( m^* \) are then used to evaluate the various causal mediation effects, which are displayed in **Output 33.2.3** and **Output 33.2.4**.
Chapter 33: The CAUSALMED Procedure

Output 33.2.3 Evaluation of Effects for the High-Motivation Group

<table>
<thead>
<tr>
<th>Summary of Effects: High-Motivation Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Total Effect</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
</tr>
<tr>
<td>Percentage Mediated</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
</tr>
</tbody>
</table>

Output 33.2.4 Evaluation of Effects for the Low-Motivation Group

<table>
<thead>
<tr>
<th>Summary of Effects: Low-Motivation Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Total Effect</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
</tr>
<tr>
<td>Percentage Mediated</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
</tr>
</tbody>
</table>

Output 33.2.3 and Output 33.2.4 show that the total effect remains the same in the two evaluations, as expected. Because the controlled direct effect is defined at a particular level of the mediator level (m*), it is not surprising that the two evaluations lead to different estimates of the controlled direct effect.

At one standard deviation above the mean of Motivation, the controlled direct effect is 4.34. This is higher than the controlled direct effect at one standard deviation below the mean, which is 4.02. The percentages of the total effect that are due to interaction also differ for the two levels of Motivation. One percentage is –2% and the other is 3%, although both are small and negligible.

You can also use the EVALUATE statement to evaluate causal mediation effects for particular target groups. The following EVALUATE statements estimate causal mediation effects for small families (FamSize=3) and large families (FamSize=7):

```r
proc causalmed data=Cognitive;
  model CogPerform = Encourage | Motivation;
  mediator Motivation = Encourage;
  covar FamSize SocStatus;
  evaluate 'Small Families' FamSize=3;
  evaluate 'Large Families' FamSize=7;
  run;
```
Output 33.2.5 and Output 33.2.6 display the corresponding effect summaries.

**Output 33.2.5** Mediation Effects Conditional on Small Families

<table>
<thead>
<tr>
<th>Summary of Effects: Small Families</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Total Effect</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
</tr>
<tr>
<td>Percentage Mediated</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
</tr>
</tbody>
</table>

**Output 33.2.6** Mediation Effects Conditional on Large Families

<table>
<thead>
<tr>
<th>Summary of Effects: Large Families</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Total Effect</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
</tr>
<tr>
<td>Percentage Mediated</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
</tr>
</tbody>
</table>

The patterns of all causal effects are similar for small families and large families. Small families appear to have a slightly higher total effect. For both groups, the percentage of the total effect that is due to the interaction between Encourage and Motivation is very small. For both groups, about 40% of the total effect is due to the mediation of Motivation.

The next set of EVALUATE statements estimate causal mediation effects for subjects whose social status (SocStatus) is high or low.

```plaintext
proc causalmed data=Cognitive;
   model CogPerform = Encourage | Motivation;
   mediator Motivation = Encourage;
   covar FamSize SocStatus;
   evaluate 'High Social Status' SocStatus=1(SD);
   evaluate 'Low Social Status' SocStatus=-1(SD);
run;
```
Output 33.2.7 and Output 33.2.8 display the corresponding effect summaries.

**Output 33.2.7** Mediation Effects Conditional on High Social Status

<table>
<thead>
<tr>
<th>Summary of Effects: High Social Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Total Effect</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
</tr>
<tr>
<td>Percentage Mediated</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
</tr>
</tbody>
</table>

**Output 33.2.8** Mediation Effects Conditional on Low Social Status

<table>
<thead>
<tr>
<th>Summary of Effects: Low Social Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Total Effect</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
</tr>
<tr>
<td>Percentage Mediated</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
</tr>
</tbody>
</table>

Again, the patterns of all causal effects are similar for both groups. The high social status group appears to have a slightly higher total effect.

You can also combine the specifications of covariates to evaluate specific causal mediation effects. In the following EVALUATE statements, subjects are defined by a combination of levels of FamSize and SocStatus:

```plaintext
proc causalmed data=Cognitive;
  model CogPerform = Encourage | Motivation;
  mediator Motivation = Encourage;
  covar FamSize SocStatus;
  evaluate 'Most Favorable Environment' FamSize=-.5(SD) SocStatus=1(SD);
  evaluate 'Least Favorable Environment' FamSize=.5(SD) SocStatus=-1(SD);
run;
```

The effects labeled ‘Most Favorable Environment’ are defined by FamSize at 0.5 standard deviation below the mean family size and SocStatus at 1 standard deviation above the mean social status rating. The effects labeled ‘Least Favorable Environment’ are defined by FamSize at 0.5 standard deviation above the mean family size and SocStatus at 1 standard deviation below the mean social status rating.
Output 33.2.9 and Output 33.2.10 display the corresponding effect summaries.

**Output 33.2.9** Mediation Effects Conditional on Most Favorable Environment

<table>
<thead>
<tr>
<th>Summary of Effects: Most Favorable Environment</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect</td>
<td>6.8969</td>
<td>0.1371</td>
<td>6.6281</td>
<td>7.1656</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876</td>
<td>4.2717</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
<td>4.2057</td>
<td>0.0476</td>
<td>4.1125</td>
<td>4.2989</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
<td>2.6912</td>
<td>0.1453</td>
<td>2.4065</td>
<td>2.9759</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>39.0202</td>
<td>1.3963</td>
<td>36.8235</td>
<td>41.0570</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>1.2102</td>
<td>0.0979</td>
<td>1.0183</td>
<td>1.4020</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>39.3078</td>
<td>1.2900</td>
<td>36.8694</td>
<td>41.0926</td>
</tr>
</tbody>
</table>

**Output 33.2.10** Mediation Effects Conditional on Least Favorable Environment

<table>
<thead>
<tr>
<th>Summary of Effects: Least Favorable Environment</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect</td>
<td>6.7874</td>
<td>0.1489</td>
<td>6.4995</td>
<td>7.0793</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
<td>4.1797</td>
<td>0.0470</td>
<td>4.0876</td>
<td>4.2717</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
<td>4.0962</td>
<td>0.0474</td>
<td>4.0033</td>
<td>4.1891</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
<td>2.6912</td>
<td>0.1453</td>
<td>2.4065</td>
<td>2.9759</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>39.6497</td>
<td>1.3438</td>
<td>37.0160</td>
<td>42.2835</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>-0.3836</td>
<td>0.1222</td>
<td>-0.6231</td>
<td>-0.1441</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>38.4201</td>
<td>1.4276</td>
<td>35.6221</td>
<td>41.2180</td>
</tr>
</tbody>
</table>

The patterns of all causal mediation effects are similar for the two groups. The total effect for ‘Most Favorable Environment’ is slightly larger than the total effect for ‘Least Favorable Environment’.

Together, these evaluations show that about 40% of the effect of parental encouragement on cognitive development is mediated by children’s learning motivation. The interaction effect of parental encouragement and children’s learning motivation is small. And more importantly, these conclusions appear to hold for different family sizes and levels of social status.

For more information about setting the covariate levels, see the **EVALUATE** statement and the section “Evaluating Causal Mediation Effects” on page 2104.

**Example 33.3: Smoking Effect on Infant Mortality**

This example demonstrates causal mediation analysis with treatment, outcome, and mediator variables that are all binary. The data contain information about infant mortality in 2003 and were obtained from the US National Center for Health Statistics. A random sample of 100,000 observations is used in this example. The analysis and its interpretation are purely illustrative; definitive conclusions should not be drawn from this example.
The following statements print the first 10 observations of the data set, which are shown in Output 33.3.1:

```plaintext
proc print data=sashelp.birthwgt(obs=10);
run;
```

**Output 33.3.1** First 10 Observations of birthwgt Data Set

**First 10 Observations of Data Set Cognitive**

<table>
<thead>
<tr>
<th>Obs</th>
<th>LowBirthWgt</th>
<th>Married</th>
<th>AgeGroup</th>
<th>Race</th>
<th>Drinking</th>
<th>Death</th>
<th>Smoking</th>
<th>SomeCollege</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
<td>No</td>
<td>3</td>
<td>Asian</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>No</td>
<td>2</td>
<td>White</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>2</td>
<td>Native</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>No</td>
<td>2</td>
<td>White</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>No</td>
<td>2</td>
<td>White</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>No</td>
<td>2</td>
<td>White</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>No</td>
<td>2</td>
<td>Asian</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>No</td>
<td>3</td>
<td>White</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Yes</td>
<td>1</td>
<td>Black</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>No</td>
<td>2</td>
<td>Native</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
```

The main variables in the analysis are as follows:

- **The treatment variable is** Smoking. It is an indicator of maternal smoking behavior, with values 'Yes' and 'No'.
- **The outcome variable is** Death. It is an indicator of infant death within one year of birth, with values 'Yes' and 'No'.
- **The mediator variable is** LowBirthWgt. It is an indicator of low birth weight (less than 2,500 grams), with values 'Yes' and 'No'.

The analysis also includes five confounding covariates:

- **AgeGroup** represents maternal ages of less than 20, between 20 and 35, and greater than 35, with values 1, 2, and 3, respectively.
- **Drinking** is an indicator of maternal drinking during pregnancy, with values 'Yes' and 'No'.
- **Married** is an indicator of marital status, with values 'Yes' and 'No'.
- **Race** is an indicator of race, with values 'Asian', 'Black', 'Hispanic', 'Native' (native American), and 'White'.
- **SomeCollege** is an indicator of whether the mother has 12 or more years of education, with values 'Yes' and 'No'.

The following statements specify a causal mediation model:
proc causalmed data=sashelp.birthwgt decomp;
  class LowBirthWgt Smoking Death AgeGroup Married Race
       Drinking SomeCollege /descending;
  mediator LowBirthWgt = Smoking;
  model Death = LowBirthWgt | Smoking;
  covar AgeGroup Married Race Drinking SomeCollege;
  evaluate 'Low Birth-Weight' LowBirthWgt='Yes' / nodecomp;
  evaluate 'Normal Birth-Weight' LowBirthWgt='No' / nodecomp;
run;

The DECOMP option requests various total effect decompositions. The MEDIATOR statement specifies
the mediator model for the response LowBirthWgt. The MODEL statement specifies the outcome model for
the response Death and assumes an interaction between LowBirthWgt and Smoking. The CLASS statement
names the categorical variables in the analysis, and the DESCENDING option models the probability of
the last level of both responses (Death='Yes' and LowBirthWgt='Yes'). The COV AR statement specifies
the five covariates. Finally, the two EVALUATE statements specify the mediator levels for comparing their
patterns of causal mediation effects.

Output 33.3.2 displays the model information, which includes the outcome, treatment, and mediator variables,
the distributions, and the link functions of the response variables. Because observations that have missing
values are not included, only 93,292 observations are used for analysis.

Output 33.3.3 displays the levels of the categorical variables, including binary response variables.
Output 33.3.4 displays frequency counts of the binary outcome, mediator, and treatment variables. It also shows which levels of the response variables are being modeled.

**Output 33.3.4** Profiles of Binary Outcome, Mediator, and Treatment Variables

<table>
<thead>
<tr>
<th>Response Profile</th>
<th>Ordered Value</th>
<th>Death Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Yes</td>
<td>527</td>
<td></td>
</tr>
<tr>
<td>2 No</td>
<td>92765</td>
<td></td>
</tr>
</tbody>
</table>

Outcome probability modeled is Death='Yes'.

<table>
<thead>
<tr>
<th>Mediator Profile</th>
<th>Ordered Value</th>
<th>LowBirthWgt</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Yes</td>
<td>7562</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 No</td>
<td>85730</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mediator probability modeled is LowBirthWgt='Yes'.

<table>
<thead>
<tr>
<th>Treatment Profile</th>
<th>Ordered Value</th>
<th>Smoking</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Yes</td>
<td>20984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 No</td>
<td>72308</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Output 33.3.5 displays the major decompositions of effects on infant mortality on both the odds ratio (OR) scale and the excess relative risk scale. Percentages of the total effect are displayed only on the excess relative risk scale.

**Output 33.3.5** Summary of Effects on Infant Mortality

<table>
<thead>
<tr>
<th>Summary of Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds Ratio Total Effect</td>
<td>1.7071</td>
<td>0.2215</td>
<td>1.2729</td>
<td>2.1412</td>
</tr>
<tr>
<td>Odds Ratio Controlled Direct Effect (CDE)</td>
<td>1.8940</td>
<td>0.3540</td>
<td>1.2002</td>
<td>2.5879</td>
</tr>
<tr>
<td>Odds Ratio Natural Direct Effect (NDE)</td>
<td>1.3626</td>
<td>0.1768</td>
<td>1.0160</td>
<td>1.7092</td>
</tr>
<tr>
<td>Odds Ratio Natural Indirect Effect (NIE)</td>
<td>1.2528</td>
<td>0.0343</td>
<td>1.1855</td>
<td>1.3201</td>
</tr>
<tr>
<td>Total Excess Relative Risk</td>
<td>0.7071</td>
<td>0.2215</td>
<td>0.2729</td>
<td>1.1412</td>
</tr>
<tr>
<td>Excess Relative Risk Due to CDE</td>
<td>0.3246</td>
<td>0.1207</td>
<td>0.0881</td>
<td>0.5611</td>
</tr>
<tr>
<td>Excess Relative Risk Due to NDE</td>
<td>0.3626</td>
<td>0.1768</td>
<td>0.0160</td>
<td>0.7092</td>
</tr>
<tr>
<td>Excess Relative Risk Due to NIE</td>
<td>0.3445</td>
<td>0.0612</td>
<td>0.2245</td>
<td>0.4644</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>48.7165</td>
<td>9.8917</td>
<td>29.3291</td>
<td>68.1040</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>81.202</td>
<td>19.8380</td>
<td>-30.7615</td>
<td>47.0020</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>54.0930</td>
<td>11.6647</td>
<td>31.2306</td>
<td>76.9554</td>
</tr>
</tbody>
</table>

The first four rows of the table in Output 33.3.5 summarize the effects on the odds ratio scale. The controlled direct effect (CDE) on this scale is 1.894. This is the CDE when the mediator variable LowBirthWgt is controlled at the level 'No'. In other words, this is the CDE odds ratio for the group that has normal birth weights. The corresponding confidence interval is (1.200, 2.588). The natural direct effect (NDE) and natural
The next seven rows of the table in Output 33.3.5 summarize effects on the excess relative risk (ERR) scale. The natural direct effect (0.363) and natural indirect effect (0.345) have an additive property on this scale; they sum to the total excess relative risk, which is 0.707. Additivity makes it easier to use these values to deduce the ‘Percentage Mediated’, which is 48.72% (= 0.3445/0.7071 × 100%). Therefore, about 50% of the smoking effect on infant mortality is mediated through the lowering of babies’ birth weights. However, the 95% confidence interval for the ‘Percentage Mediated’ is (29.3%, 68.1%), which is fairly wide. More data would yield a more precise interval estimate.

The percentage of total effect due to the interaction between smoking and low birth weights is about 8%, which is relatively small. Again, the corresponding 95% confidence interval, (–30.8%, 47.0%), is quite wide.

The DECOMP option requests various total effect decompositions, which are shown in Output 33.3.6. Following VanderWeele (2014), all these decompositions are computed on the excess relative risk scale.

### Output 33.3.6 Decompositions of Smoking Effects on Infant Mortality

| Decomposition | Excess Relative Risk | Estimate | Standard Error | Wald 95% Confidence Limits | Z Pr > |Z| |
|---------------|----------------------|----------|----------------|----------------------------|--------|-----|
| NDE+NIE       | Natural Direct       | 0.3626   | 0.1768         | 0.01604                    | 0.7092 | 2.05| 0.0403 |
|               | Natural Indirect     | 0.3445   | 0.0612         | 0.2245                     | 0.4664 | 5.63| <0.0001 |
| CDE+PE        | Controlled Direct    | 0.3246   | 0.1207         | 0.08810                    | 0.5611 | 2.69| 0.0071 |
|               | Portion Eliminated   | 0.3825   | 0.1556         | 0.07752                    | 0.6874 | 2.46| 0.0140 |
| TDE+PIE       | Total Direct         | 0.3820   | 0.2188         | -0.04688                   | 0.8109 | 1.75| 0.0809 |
|               | Pure Indirect        | 0.3251   | 0.0353         | 0.2558                     | 0.3943 | 9.20| <0.0001 |
| NDE+PIE+IMD   | Natural Direct       | 0.3626   | 0.1768         | 0.01604                    | 0.7092 | 2.05| 0.0403 |
|               | Pure Indirect        | 0.3251   | 0.0353         | 0.2558                     | 0.3943 | 9.20| <0.0001 |
|               | Mediated Interaction | 0.0194   | 0.0523         | -0.08309                   | 0.1219 | 0.37| 0.7106 |
| CDE+PIE+PAI   | Controlled Direct    | 0.3246   | 0.1207         | 0.08810                    | 0.5611 | 2.69| 0.0071 |
|               | Pure Indirect        | 0.3251   | 0.0353         | 0.2558                     | 0.3943 | 9.20| <0.0001 |
|               | Portion Due to Interaction | 0.0574 | 0.1547         | -0.2457                    | 0.3606 | 0.37| 0.7105 |
| Four-Way      | Controlled Direct    | 0.3246   | 0.1207         | 0.08810                    | 0.5611 | 2.69| 0.0071 |
|               | Reference Interaction| 0.0380   | 0.1024         | -0.1627                    | 0.2387 | 0.37| 0.7105 |
|               | Mediated Interaction | 0.0194   | 0.0523         | -0.08309                   | 0.1219 | 0.37| 0.7106 |
|               | Pure Indirect        | 0.3251   | 0.0353         | 0.2558                     | 0.3943 | 9.20| <0.0001 |
| Total         | Excess Relative Risk | 0.7071   | 0.2215         | 0.2729                     | 1.1412 | 3.19| 0.0014 |

Note: NDE=CDE+IRF, NIE=PIE+IMD, PAI=IRF+IMD, PE=PAI+PIE, TDE=CDE+PAI.

As shown in Output 33.3.7, PROC CAUSALMED also displays the corresponding decompositions by their percentage contribution to the total effect on the excess relative risk scale.
Chapter 33: The CAUSALMED Procedure

Output 33.3.7  Percentage Decomposition of Smoking Effects on Infant Mortality

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>Excess Relative Risk</th>
<th>Percent</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z  Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDE+NIE</td>
<td>Natural Direct</td>
<td>51.28</td>
<td>9.89</td>
<td>31.90</td>
<td>70.67</td>
</tr>
<tr>
<td></td>
<td>Natural Indirect</td>
<td>48.72</td>
<td>9.89</td>
<td>29.33</td>
<td>68.10</td>
</tr>
<tr>
<td>CDE+PE</td>
<td>Controlled Direct</td>
<td>45.91</td>
<td>11.66</td>
<td>23.04</td>
<td>68.77</td>
</tr>
<tr>
<td></td>
<td>Portion Eliminated</td>
<td>54.09</td>
<td>11.66</td>
<td>31.23</td>
<td>76.96</td>
</tr>
<tr>
<td>TDE+PIE</td>
<td>Total Direct</td>
<td>54.03</td>
<td>14.49</td>
<td>25.62</td>
<td>82.43</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>45.97</td>
<td>14.49</td>
<td>17.57</td>
<td>74.38</td>
</tr>
<tr>
<td>NDE+PIE+IMD</td>
<td>Natural Direct</td>
<td>51.28</td>
<td>9.89</td>
<td>31.90</td>
<td>70.67</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>45.97</td>
<td>14.49</td>
<td>17.57</td>
<td>74.38</td>
</tr>
<tr>
<td></td>
<td>Mediated Interaction</td>
<td>2.74</td>
<td>6.70</td>
<td>-10.40</td>
<td>15.88</td>
</tr>
<tr>
<td>CDE+PIE+PAI</td>
<td>Controlled Direct</td>
<td>45.91</td>
<td>11.66</td>
<td>23.04</td>
<td>68.77</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>45.97</td>
<td>14.49</td>
<td>17.57</td>
<td>74.38</td>
</tr>
<tr>
<td>Four-Way</td>
<td>Portion Due to Interaction</td>
<td>8.12</td>
<td>19.84</td>
<td>-30.76</td>
<td>47.00</td>
</tr>
<tr>
<td></td>
<td>Reference Interaction</td>
<td>5.38</td>
<td>13.14</td>
<td>-20.37</td>
<td>31.13</td>
</tr>
<tr>
<td></td>
<td>Mediated Interaction</td>
<td>2.74</td>
<td>6.70</td>
<td>-10.40</td>
<td>15.88</td>
</tr>
<tr>
<td></td>
<td>Pure Indirect</td>
<td>45.97</td>
<td>14.49</td>
<td>17.57</td>
<td>74.38</td>
</tr>
</tbody>
</table>

Note: NDE=CDE+IRF, NIE=PIE+IMD, PAI=IRF+IMD, PE=PAI+PIE, TDE=CDE+PAI.

The entries for the four-way decomposition in Output 33.3.7 show that 46% of the total effect is attributed to neither interaction nor mediation (‘Controlled Direct’), 5% is attributed to interaction but not mediation (‘Reference Interaction’), 3% is attributed to both mediation and interaction (‘Mediated Interaction’), and 46% is attributed to mediation but not interaction (‘Pure Indirect’).

In the three-way decomposition labeled ‘CDE+PIE+PAI,’ the percentage of total effect that is attributed to the interaction (PAI or ‘Portion Due to Interaction’ in the table) is about 8%, which is not large but is also not ignorable.

Note that some of the confidence intervals in this table span from negative to positive values. This indicates that the corresponding point estimates might not be very accurate.

As requested by the first EVALUATE statement, the table in Output 33.3.8 displays the major effects and percentages when the mediator LowBirthWgt is set at the level ‘Yes’.
Output 33.3.8  Summary of Smoking Effects for the Low Birth-Weight Group

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds Ratio Total Effect</td>
<td>1.7071</td>
<td>0.2215</td>
<td>1.2729</td>
<td>2.1412</td>
<td>7.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Odds Ratio Controlled Direct Effect (CDE)</td>
<td>1.0917</td>
<td>0.1591</td>
<td>0.7799</td>
<td>1.4036</td>
<td>6.86</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Odds Ratio Natural Direct Effect (NDE)</td>
<td>1.3626</td>
<td>0.1768</td>
<td>1.0160</td>
<td>1.7092</td>
<td>7.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Odds Ratio Natural Indirect Effect (NIE)</td>
<td>1.2528</td>
<td>0.0343</td>
<td>1.1855</td>
<td>1.3201</td>
<td>36.50</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Total Excess Relative Risk</td>
<td>0.7071</td>
<td>0.2215</td>
<td>0.2729</td>
<td>1.1412</td>
<td>3.19</td>
<td>0.0014</td>
</tr>
<tr>
<td>Excess Relative Risk Due to CDE</td>
<td>0.8669</td>
<td>1.4959</td>
<td>-2.0649</td>
<td>3.7988</td>
<td>0.58</td>
<td>0.5622</td>
</tr>
<tr>
<td>Excess Relative Risk Due to NDE</td>
<td>0.3626</td>
<td>0.1768</td>
<td>0.01604</td>
<td>0.7092</td>
<td>2.05</td>
<td>0.0403</td>
</tr>
<tr>
<td>Excess Relative Risk Due to NIE</td>
<td>0.3445</td>
<td>0.0612</td>
<td>0.02245</td>
<td>0.4644</td>
<td>5.63</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>48.7165</td>
<td>9.8917</td>
<td>29.3291</td>
<td>68.1040</td>
<td>4.92</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>-68.5853</td>
<td>167.6</td>
<td>-397.04</td>
<td>259.87</td>
<td>-0.41</td>
<td>0.6823</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>-22.6126</td>
<td>179.6</td>
<td>-374.60</td>
<td>329.38</td>
<td>-0.13</td>
<td>0.8998</td>
</tr>
</tbody>
</table>

The odds ratio CDE (which is evaluated for the low birth-weight group) is 1.09, with a corresponding 95% confidence interval of (0.78, 1.40).

As requested by the second EVALUATE statement, the table in Output 33.3.9 displays the major effects and percentages when the mediator LowBirthWgt is set at the level 'No'.

Output 33.3.9  Summary of Smoking Effects for the Normal Birth-Weight Group

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds Ratio Total Effect</td>
<td>1.7071</td>
<td>0.2215</td>
<td>1.2729</td>
<td>2.1412</td>
<td>7.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Odds Ratio Controlled Direct Effect (CDE)</td>
<td>1.8940</td>
<td>0.3540</td>
<td>1.2002</td>
<td>2.5879</td>
<td>5.35</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Odds Ratio Natural Direct Effect (NDE)</td>
<td>1.3626</td>
<td>0.1768</td>
<td>1.0160</td>
<td>1.7092</td>
<td>7.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Odds Ratio Natural Indirect Effect (NIE)</td>
<td>1.2528</td>
<td>0.0343</td>
<td>1.1855</td>
<td>1.3201</td>
<td>36.50</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Total Excess Relative Risk</td>
<td>0.7071</td>
<td>0.2215</td>
<td>0.2729</td>
<td>1.1412</td>
<td>3.19</td>
<td>0.0014</td>
</tr>
<tr>
<td>Excess Relative Risk Due to CDE</td>
<td>0.3246</td>
<td>0.1207</td>
<td>0.08810</td>
<td>0.5611</td>
<td>2.69</td>
<td>0.0071</td>
</tr>
<tr>
<td>Excess Relative Risk Due to NDE</td>
<td>0.3626</td>
<td>0.1768</td>
<td>0.01604</td>
<td>0.7092</td>
<td>2.05</td>
<td>0.0403</td>
</tr>
<tr>
<td>Excess Relative Risk Due to NIE</td>
<td>0.3445</td>
<td>0.0612</td>
<td>0.02245</td>
<td>0.4644</td>
<td>5.63</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>48.7165</td>
<td>9.8917</td>
<td>29.3291</td>
<td>68.1040</td>
<td>4.92</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>8.1202</td>
<td>19.8380</td>
<td>-30.7615</td>
<td>47.0020</td>
<td>0.41</td>
<td>0.6823</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>54.0930</td>
<td>11.6647</td>
<td>31.2306</td>
<td>76.9554</td>
<td>4.64</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

The odds ratio CDE (which is evaluated for the normal birth-weight group) is now 1.89, with a corresponding 95% confidence interval of (1.20, 2.59).

Note that the controlled level of the mediator requested by the second EVALUATE statement coincides the default setting that uses the last level of mediator as the controlled level. Hence, the results in Output 33.3.9 and Output 33.3.5 are identical.
Example 33.4: Mediation Analysis by Linear Structural Equation Modeling

This example illustrates the use of linear structural equation modeling and the CALIS procedure for doing a limited form of mediation analysis. For this analysis, the CALIS procedure and the CAUSALMED procedure produce results that are very similar. However, the more general approach implemented in the CAUSALMED procedure is needed to define and compute the mediation effects in a broader context. Within this context, Example 33.1 illustrates how the general approach deals with interaction effects, and Example 33.3 illustrates how it treats binary outcomes and binary mediators in a unified fashion.

The scenario in this example is the observational study that is presented in the section “Getting Started: CAUSALMED Procedure” on page 2075. The goals of the study are to determine whether an encouraging environment provided by parents (which is represented by the variable Encourage) has an effect on the cognitive development of children (which is represented by the variable CogPerform) and to estimate the amount of the total causal effect that is due to the mediation of learning motivation (which is represented by the variable Motivation).

In the example in the section “Getting Started: CAUSALMED Procedure” on page 2075, PROC CAUSALMED is used to carry out two mediation analyses:

```
proc causalmed data=Cognitive;
  model       CogPerform = Encourage Motivation;
  mediator    Motivation = Encourage;
run;
```

```
proc causalmed data=Cognitive;
  model       CogPerform = Encourage Motivation;
  mediator    Motivation = Encourage;
  covar FamSize SocStatus;
run;
```

The first analysis does not specify any confounding covariates. It produces the summary of effects in Output 33.4.1, which shows that the ‘Percentage Mediated’ is about 47%.

Output 33.4.1 Estimation of Causal Effects without Adjusting for Confounding Covariates

|                       | Estimate | Standard Error | Wald 95% Confidence Limits | Z Pr > |Z| |
|-----------------------|----------|----------------|---------------------------|--------|----|
| Total Effect          | 8.0423   | 0.0320         | 7.9796                    | 8.1050 | <.0001 |
| Controlled Direct Effect (CDE) | 4.2835 | 0.1062         | 4.0754                    | 4.4917 | 40.33 <.0001 |
| Natural Direct Effect (NDE) | 4.2835 | 0.1062         | 4.0754                    | 4.4917 | 40.33 <.0001 |
| Natural Indirect Effect (NIE) | 3.7588 | 0.1091         | 3.5449                    | 3.9727 | 34.44 <.0001 |
| Percentage Mediated   | 46.7377  | 1.3254         | 44.1400                   | 49.3353 | 35.26 <.0001 |
| Percentage Due to Interaction | 0  | .              | .                         | .      | .      |
| Percentage Eliminated  | 46.7377  | 1.3254         | 44.1400                   | 49.3353 | 35.26 <.0001 |

The second analysis specifies confounding covariates. It produces the summary of effects in Output 33.4.2. These effects have more appropriate causal interpretations if FamSize and SocStatus are the only important confounding variables that must be controlled for. Controlling for covariates, Output 33.4.2 shows a more conservative ‘Percentage Mediated’ of 37%.
Example 33.4: Mediation Analysis by Linear Structural Equation Modeling

Output 33.4.2  Estimation of Causal Effects Adjusting for Confounding Covariates

<table>
<thead>
<tr>
<th>Summary of Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Z Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect</td>
<td>6.8435</td>
<td>0.1525</td>
<td>6.5446 7.1424</td>
<td>44.88 &lt;.0001</td>
</tr>
<tr>
<td>Controlled Direct Effect (CDE)</td>
<td>4.2962</td>
<td>0.1098</td>
<td>4.0811 4.5114</td>
<td>39.14 &lt;.0001</td>
</tr>
<tr>
<td>Natural Direct Effect (NDE)</td>
<td>4.2962</td>
<td>0.1098</td>
<td>4.0811 4.5114</td>
<td>39.14 &lt;.0001</td>
</tr>
<tr>
<td>Natural Indirect Effect (NIE)</td>
<td>2.5473</td>
<td>0.1563</td>
<td>2.2410 2.8536</td>
<td>16.30 &lt;.0001</td>
</tr>
<tr>
<td>Percentage Mediated</td>
<td>37.2219</td>
<td>1.7523</td>
<td>33.7874 40.6564</td>
<td>21.24 &lt;.0001</td>
</tr>
<tr>
<td>Percentage Due to Interaction</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Percentage Eliminated</td>
<td>37.2219</td>
<td>1.7523</td>
<td>33.7874 40.6564</td>
<td>21.24 &lt;.0001</td>
</tr>
</tbody>
</table>

The second analysis decomposes the total effect of an encouraging environment on cognitive development into two percentages:

- 63% of the total effect is through the direct pathway Encourage→CogPerform
- 37% of the total effect is through the mediation pathway Encourage→Motivation→CogPerform

Statements such as these invite the use of structural equation modeling, which offers the same type of language for describing causal sequences. Indeed, mediation analysis has a relatively long history in the field of psychology, where structural equation modeling is quite popular.

By specifying the relevant causal pathways in structural equation models, you can use the CALIS procedure to obtain essentially the same mediation analyses as those obtained with the CAUSALMED procedure:

```latex
proc calis data=cognitive;
  path
    Encourage ===> Motivation,
    Encourage Motivation ===> CogPerform;
  effpart Encourage ===> CogPerform;
run;
```

```latex
proc calis data=cognitive;
  path
    Encourage ===> Motivation,
    Encourage Motivation ===> CogPerform,
    FamSize ===> Encourage Motivation CogPerform,
    SocStatus ===> Encourage Motivation CogPerform;
  effpart Encourage ===> CogPerform;
run;
```

The EFFPART statements request the total effect decompositions of Encourage on CogPerform in the two analyses. Output 33.4.3 shows the total effect decomposition when the covariates are ignored in the linear structural equation model. The total, direct, and indirect effects and their standard error estimates closely match those in Output 33.4.1.
Output 33.4.3 Summary of Causal Effects

<table>
<thead>
<tr>
<th>Effects of Encourage</th>
<th>Total Effect</th>
<th>Direct Effect</th>
<th>Indirect Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>CogPerform</td>
<td>8.0423</td>
<td>4.2835</td>
<td>3.7588</td>
</tr>
<tr>
<td></td>
<td>0.0321</td>
<td>0.1064</td>
<td>0.1093</td>
</tr>
<tr>
<td></td>
<td>250.8761</td>
<td>40.2662</td>
<td>34.3806</td>
</tr>
<tr>
<td></td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Output 33.4.4 shows the total effect decomposition when the covariates are incorporated in the linear structural equation model. Again, the total, direct, and indirect effects and their standard error estimates closely match those in Output 33.4.2.

Output 33.4.4 Summary of Causal Effects

<table>
<thead>
<tr>
<th>Effects of Encourage</th>
<th>Total Effect</th>
<th>Direct Effect</th>
<th>Indirect Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>CogPerform</td>
<td>6.8435</td>
<td>4.2962</td>
<td>2.5473</td>
</tr>
<tr>
<td></td>
<td>0.1527</td>
<td>0.1099</td>
<td>0.1565</td>
</tr>
<tr>
<td></td>
<td>44.8044</td>
<td>39.0749</td>
<td>16.2739</td>
</tr>
<tr>
<td></td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

However, the similarity of the analyses obtained with the CALIS and CAUSALMED procedures does not extend to more general situations. The limitations of structural equation modeling include the following:

- It does not have a clear foundation for defining causal mediation effects.
- It does not deal with interaction effects effectively.
- It does not treat binary outcomes and binary mediators in a unified fashion.

The general mediation approach that is implemented in PROC CAUSALMED overcomes these limitations of traditional linear structural equation modeling. For more information about the theoretical foundation of the general mediation approach, see the section “Causal Mediation Effects: Definitions, Assumptions, and Identification” on page 2097.

References


Subject Index

bias-corrected confidence interval
  CAUSALMED procedure, 2087, 2112
bootstrap normal confidence interval
  CAUSALMED procedure, 2111
bootstrap percentile method
  CAUSALMED procedure, 2087

CAUSALMED procedure
  bias-corrected confidence interval, 2087, 2112
  bootstrap normal confidence interval, 2111
  bootstrap percentile method, 2087
  conditional effect evaluation, 2091
  confidence interval, 2087
  mediator model, 2095
  ODS table names, 2112
  output table names, 2112
  percentile confidence interval, 2111
  printing options, 2112
classification variables
  sort order of levels (CAUSALMED), 2085
conditional effect evaluation
  CAUSALMED procedure, 2091
  confidence interval
    CAUSALMED procedure, 2087
mediator model
  CAUSALMED procedure, 2095
output table names
  CAUSALMED procedure, 2112
percentile confidence interval
  CAUSALMED procedure, 2111
probability distribution, built-in
  CAUSALMED procedure, 2096
Syntax Index

ABSCONV option
  PROC CAUSALMED statement, 2083
ABSFCONV option
  PROC CAUSALMED statement, 2083
ABSGCONV option
  PROC CAUSALMED statement, 2083
ALPHA= option
  PROC CAUSALMED statement, 2081

BOOTSTRAP statement
  CAUSALMED procedure, 2087
BY statement
  CAUSALMED procedure, 2088

CAUSALMED procedure, 2080
  syntax, 2080
CAUSALMED procedure, BOOTSTRAP statement, 2087
  BOOTCI option, 2087
  NBOOT option, 2088
  SEED option, 2088
CAUSALMED procedure, BY statement, 2088
CAUSALMED procedure, CLASS statement, 2089
  CPREFIX= option, 2089
  DESCENDING option, 2089
  LPREFIX= option, 2089
  MISSING option, 2089
  ORDER= option, 2089
  REF= option, 2090
  TRUNCATE option, 2090
CAUSALMED procedure, COVAR statement, 2090
CAUSALMED procedure, EVALUATE statement, 2091
CAUSALMED procedure, FREQ statement, 2094
CAUSALMED procedure, MEDIATOR statement, 2095
CAUSALMED procedure, MODEL statement, 2095
  DIST= option, 2096
  DISTRIBUTION= option, 2096
  LINK= option, 2096
CAUSALMED procedure, PROC CAUSALMED statement, 2081
  ABSCONV= option, 2083
  ABSFCONV= option, 2083
  ABSGCONV= option, 2083
  ALPHA= option, 2081
  DATA= option, 2082
  DECOMP option, 2082
  DESCENDING option, 2082, 2083
  FCONV= option, 2083
  GCONV= option, 2084
  MAXFUNC= option, 2084
  MAXITER= option, 2084
  MAXTIME= option, 2084
  NAMELENGTH= option, 2083
  NLOPTIONS option, 2083
  NOPRINT option, 2085
  ORDER= option, 2085
  PALL option, 2085
  PMEDMOD option, 2086
  POUTCOMEMOD option, 2086
  PSHORT option, 2086
  PSUMMARY option, 2086
  RORDER= option, 2086
  SINGULAR= option, 2086
  TECHNIQUE= option, 2085
  THREADS= option, 2087
CI option
  BOOTSTRAP statement (CAUSALMED), 2087
CLASS statement
  CAUSALMED procedure, 2089
COVAR statement
  CAUSALMED procedure, 2090
CPREFIX= option
  CLASS statement (CAUSALMED), 2089

DATA= option
  PROC CAUSALMED statement, 2082
DECOMP option
  PROC CAUSALMED statement, 2082
DESCENDING option
  CLASS statement (CAUSALMED), 2089
  PROC CAUSALMED statement, 2082, 2083
DIST= option
  MODEL statement, 2096
DISTRIBUTION= option
  MODEL statement, 2096
EVALUATE statement
  CAUSALMED procedure, 2091
FCONV option
  PROC CAUSALMED statement, 2083
FREQ statement
  CAUSALMED procedure, 2094
GCONV option
  PROC CAUSALMED statement, 2084