Chapter 36  
The CAUSALTRT Procedure  

Overview: CAUSALTRT Procedure  

The CAUSALTRT procedure estimates the average causal effect of a binary treatment, \( T \), on a continuous or discrete outcome, \( Y \). Although the causal effect that is defined and estimated in PROC CAUSALTRT is called a treatment effect, it is not confined to effects that result from controllable treatments (such as effects in an experiment). Depending on the application, the binary treatment variable \( T \) can represent an
intervention (such as smoking cessation versus control), an exposure to a condition (such as attending a private versus public school), or an existing characteristic of subjects (such as high versus low socioeconomic status). The CAUSALTRT procedure can estimate two types of causal effects: the average treatment effect (ATE) and the average treatment effect for the treated (ATT). For more information about the causal effects that the CAUSALTRT procedure can estimate, see the section “Causal Effects: Definitions, Assumptions, and Identification” on page 2394.

The CAUSALTRT procedure implements causal inference methods that are designed primarily for use with data from nonrandomized trials or observational studies. In an observational study, you observe the treatment $T$ and the outcome $Y$ without assigning subjects randomly to the treatment conditions. Instead, subjects “select” themselves into the treatment conditions according to their pretreatment characteristics. If these pretreatment characteristics are also associated with the outcome $Y$, they induce a spurious relationship between $T$ and $Y$ and hence cloud the causal interpretation of $T$ on $Y$. Therefore, estimating the causal effect of $T$ in observational studies usually requires adjustments that remove or counter the spurious effects that are induced by the confounding variables.

To adjust for the effects of the confounding variables, you can model either the treatment assignment $T$ or the outcome $Y$, or both. Modeling the treatment leads to inverse probability weighting methods, and modeling the outcome leads to regression adjustment methods. Combined modeling of the treatment and outcome leads to doubly robust methods that can provide unbiased estimates for the treatment effect even if one of the models is misspecified. For more information about which model specifications are required for different estimation methods and how default estimation methods are determined, see the section “Outline of Estimation Method Requirements” on page 2367. For an introduction to causal inference from nonrandomized data, see Imbens and Rubin (2015); Morgan and Winship (2015); Berzuini, Dawid, and Bernardinelli (2012).

---

### Features of the CAUSALTRT Procedure

The CAUSALTRT procedure provides the following methods to estimate causal effects:

- inverse probability weighting methods
- regression adjustment
- doubly robust methods

You can estimate two types of causal effects:

- average treatment effect (ATE), also sometimes called the average causal effect (ACE)
- average treatment effect for the treated (ATT or ATET)

For more information about the causal effects that the CAUSALTRT procedure can estimate, see the section “Causal Effects: Definitions, Assumptions, and Identification” on page 2394.

PROC CAUSALTRT computes standard errors and confidence intervals for the causal effects by the following methods:

- asymptotic methods
• bootstrap methods (which you request by specifying the `BOOTSTRAP` statement)

For more information about how the `CAUSALTRT` procedure computes standard errors and confidence limits, see the section “Standard Errors and Confidence Intervals” on page 2404.

PROC `CAUSALTRT` provides the following types of graphical output:

• diagnostic plots for the propensity score model, including various plots of propensity scores or weights
• histograms for bootstrap estimates

Finally, you can save the propensity scores, inverse probability weights, and the predicted potential outcomes in a SAS data set.

Outline of Estimation Method Requirements

The `CAUSALTRT` procedure uses various estimation methods to adjust for the effects of confounding variables by fitting models for the treatment assignment $T$ or the outcome $Y$, or both. For the estimation methods, you specify the outcome variable $Y$ in the `MODEL` statement and the treatment assignment $T$ in the `PSMODEL` statement.

Depending on the estimation method that you specify in the `METHOD=` option in the `PROC CAUSALTRT` statement, you must also provide specific additional modeling information in the `MODEL` and `PSMODEL` statements, as shown in Table 36.1.

<table>
<thead>
<tr>
<th><code>METHOD=</code></th>
<th>Additional Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPW, IPWR, or IPWS REGADJ</td>
<td>A model for the treatment assignment in the <code>PSMODEL</code> statement. A model for the outcome in the <code>MODEL</code> statement. If you do not specify an outcome model, PROC <code>CAUSALTRT</code> fits an intercept-only model separately for each treatment condition.</td>
</tr>
<tr>
<td>AIPW or IPWREG</td>
<td>Both an outcome model in the <code>MODEL</code> statement and a model for the treatment assignment in the <code>PSMODEL</code> statement.</td>
</tr>
</tbody>
</table>

The model for the treatment assignment is also called the propensity score model.

If you do not specify the `METHOD=` option in the `PROC CAUSALTRT` statement, the `CAUSALTRT` procedure determines a default estimation method based on which models are specified in the `MODEL` and `PSMODEL` statements, according to following criteria:

• When only a treatment model is specified (in the `PSMODEL` statement), the default estimation method is IPWR. For example:
When only an outcome model is specified (in the MODEL statement), the default estimation method is REGADJ. For example:

```
proc causaltrt;
  model y = x1 x2;
  psmodel trt;
run;
```

When both the treatment and outcome models are specified, the default estimation method is AIPW. For example:

```
proc causaltrt;
  model y = x1 x2;
  psmodel trt = x1 x2;
run;
```

When neither an outcome nor a treatment model is specified, the default estimation method is REGADJ with an intercept-only outcome model. For example:

```
proc causaltrt;
  model y;
  psmodel trt;
run;
```

If you specify the ATT option in the PROC CAUSALTRT statement, then you are estimating the average treatment effect for the treated (ATT) instead of the default average treatment effect (ATE). In this case, estimation methods are limited to either IPWR or regression adjustment. Therefore, if models for both the treatment assignment and outcome variable are specified without the METHOD=IPWR or METHOD=REGADJ option, PROC CAUSALTRT issues an error.

The CAUSALTRT procedure fits only the models necessary for the estimation method used. For example, if a model for the treatment assignment is specified but the REGADJ estimation method is requested, then PROC CAUSALTRT does not fit the treatment model and therefore does not produce output or displays that involve the propensity scores or weights.

For more information about the estimation methods that the CAUSALTRT procedure implements, see the sections “Estimating the Average Treatment Effect (ATE)” on page 2397 and “Estimating the Average Treatment Effect for the Treated (ATT)” on page 2402.
Getting Started: CAUSALTRT Procedure

This section illustrates some of the basic features of the CAUSALTRT procedure. This example uses data from a hypothetical nonrandomized trial.

Suppose that 486 patients in a trial are at risk for developing type 2 diabetes and are allowed to choose which of two preventive drugs they want to receive. For this study, the outcome of interest is whether a patient develops type 2 diabetes within five years. The nonrandom assignment of patients to treatment conditions does not control for confounding variables that might explain any observed difference between the treatment conditions. The hypothetical data set Drugs contains the following variables:

- **Age**: age at the start of the trial
- **BMI**: body mass index at the start of the trial
- **Diabetes2**: indicator for whether a patient developed type 2 diabetes, with values Yes and No
- **Drug**: indicator of treatment assignment, with values Drug_A and Drug_X
- **Gender**: gender

The first 10 observations in the Drugs data set are listed in Figure 36.1.

![Figure 36.1 Drug Trial Data](image)

The following statements invoke the CAUSALTRT procedure and request the estimation of the average treatment effect (ATE) by using the inverse probability weighting method with ratio adjustment (METHOD=IPWR):

```r
proc causaltrt data=drugs method=ipwr ppsmodel;
  class Gender;
  psmodel Drug(ref='Drug_A') = Age Gender BMI;
  model Diabetes2(ref='No') / dist = bin;
run;
```

The results of this analysis are shown in the following figures. In Figure 36.2, PROC CAUSALTRT displays information about the estimation method used, the outcome variable, and the treatment variable.
The METHOD= option in the PROC CAUSALTRT statement specifies the estimation method to be used. The outcome variable, Diabetes2, is specified in the MODEL statement, and the treatment variable, Drug, is specified in the PSMODEL statement.

The DIST=BIN option in the MODEL statement identifies the outcome as a binary variable, and the ref='No' option identifies No as the reference level. In the PSMODEL statement, the ref='Drug_A' option identifies Drug_A as the reference (control) condition when the treatment assignment is modeled and the ATE is estimated. The frequencies for the outcome and treatment variable categories are listed in the “Response Profile” (Figure 36.3) and “Treatment Profile” (Figure 36.4) tables, respectively.

The IPWR estimation method uses inverse probability weighting to estimate the potential outcome means and ATE. The inverse probability weights are estimated by inverting the predicted probability of receiving treatment (that is, the Drug_X condition), which is estimated from a logistic regression model that is specified in the PSMODEL statement. The probability of receiving treatment is called the propensity score, and hence the regression model fit by the PSMODEL statement is also called the propensity score model. In this example, the predictors of the propensity score model are Age, Gender, and BMI. The PPSMODEL option in the PROC CAUSALTRT statement displays the parameter estimates for the propensity score model, as shown in Figure 36.5.
Figure 36.5 Propensity Score Model Estimates

The CAUSALTRT Procedure

<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>-0.6226</td>
<td>1.2738</td>
<td>-3.1193 - 1.8740</td>
<td>0.2389</td>
<td>0.6250</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0888</td>
<td>0.0170</td>
<td>-0.1220 - 0.0556</td>
<td>27.4228</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Gender Female</td>
<td>-0.2103</td>
<td>0.2055</td>
<td>-0.6130 - 0.1924</td>
<td>1.0474</td>
<td>0.3061</td>
</tr>
<tr>
<td>Gender Male</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>0.1434</td>
<td>0.0517</td>
<td>0.0422 - 0.2447</td>
<td>7.7115</td>
<td>0.0055</td>
</tr>
</tbody>
</table>

The estimates for the potential outcome means (POM) and ATE are displayed in the “Analysis of Causal Effect” table (Figure 36.6). The ATE estimate of –0.1787 indicates that Drug_X is more effective on average than Drug_A in preventing the development of type 2 diabetes. The ATE is significantly different from 0 at the 0.05 \( \alpha \)-level, as indicated by the 95% confidence interval (–0.2771, –0.0803) or the \( p \)-value (0.0004).

Figure 36.6 IPWR Potential Outcome Means and ATE Estimates

The ATE of Drug_X compared to Drug_A can also be estimated by using augmented inverse probability weights (AIPW). In addition to a propensity score model, the AIPW estimation method incorporates a model for the outcome variable, Diabetes2, into the estimation of the potential outcome means and ATE. The AIPW estimation method is doubly robust and provides unbiased estimates for the ATE even if one of the outcome or treatment models is misspecified.

The following statements invoke the CAUSALTRT procedure and use the AIPW estimation method to estimate the ATE:

```plaintext
proc causaltrt data=drugs method=aipw;
  class Gender;
  psmodel Drug(ref='Drug_A') = Age Gender BMI;
  model Diabetes2(ref='No') = Age Gender BMI / dist = bin;
run;
```

For this example, the same set of effects is specified in both the MODEL and PSMODEL statements, but in general this need not be the case. The use of the AIPW estimation method for this PROC CAUSALTRT step is reflected in the “Model Information” table (Figure 36.7).
As shown in Figure 36.8, the AIPW estimate for the ATE is –0.1709, which is similar to the IPWR estimate of –0.1787. Moreover, the AIPW standard error estimate for ATE is slightly smaller than that of the IPWR. The AIPW 95% confidence interval for ATE is also slightly narrower than that of the IPWR.

### Syntax: CAUSALTRT Procedure

The following statements are available in the CAUSALTRT procedure. Items within < > are optional.

```
PROC CAUSALTRT < options > ;
    BOOTSTRAP < options > ;
    BY variables ;
    CLASS variables <(options)> < ... < variable<(options)> > < / global-options > ;
    FREQ variable ;
    MODEL outcome < (variable-options) >= < effects > < / model-options > ;
    OUTPUT < OUT=SAS-data-set > < keyword=name . . . keyword=name > ;
    PSMODEL treatment < (variable-options) >= < effects < / psmodel-options > > ;
```

The PROC CAUSALTRT, MODEL, and PSMODEL statements are required. The CLASS statement, if present, must precede the MODEL and PSMODEL statements. All statements can appear only once.

All the estimation methods implemented in PROC CAUSALTRT require the specification of both an outcome and a treatment variable. You specify the outcome variable in the MODEL statement and the treatment variable in the PSMODEL statement. In the MODEL and PSMODEL statements, you can use the same syntax to specify main effects and interaction terms for modeling the outcome and treatment assignment. This syntax is described in the section “Specification of Effects” on page 4020 in Chapter 50, “The GLM Procedure.” For more information about which model specifications are required for different estimation
methods and how default estimation methods are determined, see the section “Outline of Estimation Method Requirements” on page 2367. The following sections describe the PROC CAUSALTRT statement and then describe the other statements in alphabetical order.

PROC CAUSALTRT Statement

PROC CAUSALTRT <options> ;

The PROC CAUSALTRT statement invokes the CAUSALTRT procedure. Table 36.2 summarizes the options available in the PROC CAUSALTRT statement.

Table 36.2 PROC CAUSALTRT Statement Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>DATA=</td>
<td>Specifies the input data set</td>
</tr>
<tr>
<td>Response and Classification Variable Options</td>
<td></td>
</tr>
<tr>
<td>DESCENDING</td>
<td>Sorts the outcome variable in reverse of the default order</td>
</tr>
<tr>
<td>NAMELEN=</td>
<td>Specifies the length of effect names</td>
</tr>
<tr>
<td>ORDER=</td>
<td>Specifies the sort order of the classification variables</td>
</tr>
<tr>
<td>RORDER=</td>
<td>Specifies the sort order of the outcome variable</td>
</tr>
<tr>
<td>Estimation and Analysis</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>Estimates the average treatment effect for the treated</td>
</tr>
<tr>
<td>COVDIFFPS</td>
<td>Computes and displays standardized mean differences for the effects in the propensity score model</td>
</tr>
<tr>
<td>METHOD=</td>
<td>Specifies the estimation method</td>
</tr>
<tr>
<td>Displayed Output</td>
<td></td>
</tr>
<tr>
<td>ALPHA=</td>
<td>Specifies the level for confidence limits</td>
</tr>
<tr>
<td>NOEFFECT</td>
<td>Suppresses all displayed output that involves the outcome variable</td>
</tr>
<tr>
<td>NOPRINT</td>
<td>Suppresses all displayed output</td>
</tr>
<tr>
<td>PALL</td>
<td>Displays all optional output</td>
</tr>
<tr>
<td>PLOTS</td>
<td>Produces ODS Graphics displays</td>
</tr>
<tr>
<td>POUTCOMEMOD</td>
<td>Displays outcome model parameter estimates</td>
</tr>
<tr>
<td>PPSMODEL</td>
<td>Displays propensity score model parameter estimates</td>
</tr>
<tr>
<td>Technical Details</td>
<td></td>
</tr>
<tr>
<td>NLOPTIONS</td>
<td>Specifies optimization parameters for fitting the specified models</td>
</tr>
<tr>
<td>SINGULAR=</td>
<td>Specifies the singularity tolerance</td>
</tr>
<tr>
<td>THREADS=</td>
<td>Specifies the number of threads for the computation</td>
</tr>
</tbody>
</table>

You can specify the following options.

**ALPHA=number**

specifies a number to be used as the $\alpha$ level for $100(1 - \alpha)\%$ confidence limits in the “Analysis of Causal Effect” table. The number must be between 0 and 1. This number is also used as the default level for propensity score and outcome model confidence limits. For the propensity score and outcome models, you can override this default level by specifying the ALPHA= options in the MODEL and
PSMODEL statements, respectively. By default, ALPHA=0.05, which results in 95% confidence intervals.

ATT
ATET

estimates the average treatment effect for the treated. When this option is applied, it replaces the default estimation of the average treatment effect (ATE). For more information about the estimation methods implemented in the CAUSALTRT procedure and for comparisons between the average treatment effect and average treatment effect for the treated, see the sections “Estimating the Average Treatment Effect for the Treated (ATT)” on page 2402 and “Causal Effects: Definitions, Assumptions, and Identification” on page 2394. This option can be used only when METHOD=IPWR or REGADJ.

COVDIFFPS

computes weighted and unweighted standardized mean differences (between treatment and control conditions) and variance ratios (treatment to control) for the covariates (effects) in the propensity score model. This option is supported only for estimation methods that fit a propensity score model.

The results are displayed in the “Covariate Differences for Propensity Score Model” table. This table also includes columns for the weighted and unweighted mean and variance for propensity score model effects within each treatment condition; these columns are not displayed but are accessible if you save the table as an output data set by specifying the ODS OUTPUT statement. You can display these columns by modifying the corresponding template.

DATA=SAS-data-set

names the SAS data set that contains the data to be analyzed. If you omit this option, the procedure uses the most recently created SAS data set.

DESCENDING
DESCEND
DESC

sorts the levels of the outcome variable for a binary model in reverse of the specified order.

METHOD= AIPW | IPW | IPWR | IPWS | REGADJ | IPWREG

specifies the method to use to estimate the potential outcomes and treatment effect. You can specify one of the following values:

AIPW

performs a doubly robust estimation by using augmented inverse probability weighting. You must specify a model for the outcome variable in the MODEL statement and a model for the treatment assignment in the PSMODEL statement.

IPW

uses a basic inverse probability weighting method. You must specify a model for the treatment assignment in the PSMODEL statement.

IPWR

uses an inverse probability weighting method with ratio adjustment. You must specify a model for the treatment assignment in the PSMODEL statement.

IPWS

uses an inverse probability weighting method with ratio and scale adjustments. You must specify a model for the treatment assignment in the PSMODEL statement.
IPWREG performs a doubly robust estimation by using inverse probability weighted regression adjustment. You must specify a model for the outcome variable in the MODEL statement and a model for the treatment assignment in the PSMODEL statement.

REGADJ uses regression adjustment. You must specify a model for the outcome variable in the MODEL statement.

For all estimation methods, you specify the outcome variable in the MODEL statement and the treatment variable in the PSMODEL statement. For more information about the estimation methods that the CAUSALTRT procedure implements, see the sections “Estimating the Average Treatment Effect (ATE)” on page 2397 and “Estimating the Average Treatment Effect for the Treated (ATT)” on page 2402.

NAMELEN=n specifies the maximum length of effect names in tables and output data sets 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

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.
Chapter 36: The CAUSALTRT Procedure

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, \( \text{FCONV}=10^{-\text{FDIGITS}} \), where by default \( \text{FDIGITS} \) 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,

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

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:

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

By default, \( \text{GCONV}=1\times10^{-8} \).

MAXFUNC=n
MAXFU=n

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

- 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 are
as follows, depending on the value of the TECHNIQUE= suboption:

- 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 from 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=NEWRAP.

For more information about these optimization methods, see the section “Choosing an Optimization Algorithm” on page 512 in Chapter 19, “Shared Concepts and Topics.”

NOEFFECT

suppresses the display of all output that involves the outcome variable. This option is useful for investigating the balance of covariates between treatment conditions by exploring different propensity score models before displaying estimates for the causal effect. This option is effective only when METHOD=IPW, IPWR, or IPWS.

NOPRINT

suppresses all displayed output. This option temporarily disables the Output Delivery System (ODS). For more information, see Chapter 20, “Using the Output Delivery System.”

ORDER=DATA | FORMATTED | FREQ | INTERNAL

specifies the sort order for the levels of the classification variables (which are specified in the CLASS statement).

This option applies to the levels for all classification variables, except when you use the (default) ORDER=FORMATTED option with numeric classification variables that have no explicit format. In that case, the levels of such variables are ordered by their internal value.

The ORDER= option can take the following values:
Value of ORDER= Levels Sorted By

<table>
<thead>
<tr>
<th>ORDER=</th>
<th>Description</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 with no explicit format, which are sorted by their unformatted (internal) value</td>
</tr>
<tr>
<td>FREQ</td>
<td>Descending frequency count; levels with the most observations come first in the order</td>
</tr>
<tr>
<td>INTERNAL</td>
<td>Unformatted value</td>
</tr>
</tbody>
</table>

By default, ORDER=FORMATTED. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent.

For more information about sort order, see the chapter on the SORT procedure in the Base SAS Procedures Guide and the discussion of BY-group processing in SAS Language Reference: Concepts.

PALL
ALL
displays all optional output.

PLOTS<(global-plot-options)>=plot-request>

controls plots that are produced through ODS Graphics. For more information about controlling specific plots, see also the PLOTS options in the BOOTSTRAP and PSMODEL statements.

ODS Graphics must be enabled before plots can be requested. For example:

    ods graphics on;
    proc causaltrt plots=all method=ipwr;
      model y;
      psmodel trt = x1 x2;
    run;
    ods graphics off;

For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 623 in Chapter 21, “Statistical Graphics Using ODS.”

You can specify the following plot-requests:

ALL
produces all plots that are available for the specified estimation method.

BOOTHIST
produces histograms of the bootstrap estimates for the potential outcome means and treatment effect, which are displayed in a panel by default. This option is ignored if the BOOTSTRAP statement is not specified.
LOGITPSCORE

LPS
produces overlaid density plots for the logit of the propensity score within each treatment condition. These plots are not produced when METHOD=REGADJ.

NONE
suppresses all plots. If you specify this plot-request, then all plot requests that are specified in the BOOTSTRAP and PSMODEL statements are ignored.

OUTBYPSCORE

OUTBYPS
produces a scatter plot of the outcome variable by the propensity score for nonbinary outcomes. If the outcome is binary, box plots of the propensity scores within treatment conditions are produced for each outcome level. The whisker lengths for each box plot are determined by the maximum and minimum of the propensity scores. This plot is not produced when METHOD=REGADJ.

OUTBYWEIGHT

OUTBYWGT
produces a scatter plot of the outcome variable by weight for nonbinary outcomes. If the outcome is binary, box plots of the weights within treatment conditions are produced for each outcome level. The whisker lengths for each box plot are determined by the maximum and minimum of the weights. This plot is not produced when METHOD=REGADJ.

PS CLOUD

produces a point cloud of the propensity scores by jittering within the control and treatment conditions. This plot is not produced when METHOD=REGADJ.

PSCOVDEN

produces density plots for the covariates or continuous effects that are specified in the PSMODEL statement. Each plot displays the density of an effect for the treatment and control conditions. Two plots are produced for each effect: one plot displays unweighted densities, and the other plot displays densities that are weighted by inverse probability weights. By default, plots are produced for all continuous effects in the PSMODEL statement and are collected in panels. You can customize the density plots by using the PLOTS= option in the PSMODEL statement. This option is ignored when METHOD=REGADJ.

PSDIST

produces a box plot of the propensity score for each treatment condition. The whisker lengths for each box plot are determined by the maximum and minimum of the propensity scores. This plot is not produced when METHOD=REGADJ.

WEIGHTCLOUD

W CLOUD
produces a point cloud of the weights by jittering within the control and treatment conditions. This plot is not produced when METHOD=REGADJ.
WEIGHTDIST
WDIST
produces a box plot of the weights for each treatment condition. The whisker lengths for each box plot are determined by the maximum and minimum of the weights. This plot is not produced when METHOD=REGADJ.

You can specify the following global-plot-options:

UNPACK
UNPACKPANEL
suppresses paneling. By default, multiple plots can appear in the same output panel. You can use this option to display each plot separately.

POUTCOMEMOD
PREGADJ
displays parameter estimates for the outcome models for the control and treatment conditions.

PPSMODEL
PTREATMOD
displays parameter estimates for the propensity score model.

RORDER=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 CAUSALTRT 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 CAUSALTRT 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. By default, tolerance is 1E7 times the machine epsilon.
**THREADS=n**

specifies the number of threads 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

```
BOOTSTRAP < options > ;
```

A BOOTSTRAP statement requests bootstrap estimation of confidence intervals for the potential outcome means and treatment effect estimates. The bootstrap samples are taken from within the treatment conditions. To produce the plots that are controlled by the BOOTSTRAP statement, ODS Graphics must be enabled. PROC CAUSALTRT ignores the BOOTSTRAP statement if the initial estimation of the treatment effect results in an error or if the NOEFFECT option is specified in the PROC CAUSALTRT statement.

Table 36.3 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 of the estimates</td>
</tr>
<tr>
<td>BOOTDATA=</td>
<td>Specifies the bootstrap output data set</td>
</tr>
<tr>
<td>NBOOT=</td>
<td>Specifies the number of bootstrap sample data sets (replicates)</td>
</tr>
<tr>
<td>NOSKIP</td>
<td>Includes bootstrap samples that have inconsistent class levels</td>
</tr>
<tr>
<td>PLOTS=</td>
<td>Produces plots of the bootstrap estimates</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)>**

computes bootstrap-based confidence intervals for the potential outcome means and the treatment effect estimates, and displays the values in the “Analysis of Causal Effect” 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:

- **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.

**ALL** produces all three confidence intervals.

The **ALPHA=** option in the **PROC CAUSALTRT** statement sets the level of significance that is used to construct the bootstrap confidence intervals. For more information about how the bootstrap-based confidence intervals are computed, see the section “Bootstrap Methods” on page 2406. By default, **BOOTCI=BC** and **PROC CAUSALTRT** produces bias-corrected confidence intervals based on 1,000 bootstrap samples.

**BOOTDATA=SAS-data-set** specifies the **SAS-data-set** that contains the estimates for the potential outcome means and treatment effect for each bootstrap sample that converges.

**NBOOT=n**

**NSAMPLE=n**

**NSAMPLES=n** specifies the number of bootstrap sample data sets (replicates). A maximum of 10,000 bootstrap sample data sets can be requested. By default, **NBOOT=1000**.

**NOSKIP** includes in the bootstrap estimation bootstrap samples that have inconsistent class levels in either the treatment or outcome model. By default, for any classification variable in a bootstrap sample data set that does not contain all levels that have been used in fitting either the treatment or outcome model for the original input data set, **PROC CAUSALTRT** treats the corresponding bootstrap sample estimates as nonconvergent and skips the estimation of the treatment effect. This option overrides the default and continues the estimation of the treatment effect for bootstrap samples that have inconsistent class levels.

**PLOTS < (global-plot-options) >= plot-request< (options) > >** produces ODS graphics of bootstrap estimates. By default, if you specify the **PLOTS** option, a panel is produced with the histograms of the bootstrap estimates for the potential outcome means and treatment effect. No graphics are produced if you specify the **PLOTS=NONE** option in the **PROC CAUSALTRT** statement.

ODS Graphics must be enabled before plots can be requested. For example:

```sas
ods graphics on;

proc causaltrt method=ipwr;
  model y;
  psmodel trt = x1 x2;
  bootstrap plot;
run;
```
ods graphics off;

For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 623 in Chapter 21, “Statistical Graphics Using ODS.”

You can specify the following plot-requests:

**HIST< (EFFECT | POMCNT | POMTRT) >**
**HISTOGRAM< (EFFECT | POMCNT | POMTRT) >**
produces histograms of the bootstrap estimates for the potential outcome means and treatment effect, which are displayed in a panel by default. You can request output of specific histograms by specifying one or more the following additional suboptions:

**EFFECT** produces a histogram of the bootstrap estimates for the treatment effect.
**POMCNT** produces a histogram of the bootstrap estimates for the potential outcome mean of the control condition.
**POMTRT** produces a histogram of the bootstrap estimates for the potential outcome mean of the treatment condition.

If you specify the ATT option in the PROC CAUSALTRT statement, then the histogram produced by the EFFECT suboption is the average treatment effect for the treated, and the histograms produced by the POMCNT and POMTRT suboptions are the potential outcome means conditioned on having received treatment. For more information about the differences between the average treatment effect and average treatment effect for the treated, see the section “Causal Effects: Definitions, Assumptions, and Identification” on page 2394.

**NONE** suppresses all plots that are usually produced by the BOOTSTRAP statement. You can use this option to suppress output if the PLOTS=ALL option is specified in the PROC CAUSALTRT statement.

You can specify the following global-plot-option:

**UNPACK**
**UNPACKPANEL**
suppresses paneling. By default, multiple histograms of bootstrap estimates appear in the same output panel. You can use this option to display each histogram separately.

**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 CAUSALTRT 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 use the same SEED=
value and the same number of threads for the analytic computations. When bootstrap resampling is
performed, a column that indicates the number of threads used is added to the “Analysis of Causal
Effect” table; this column is not displayed but is available if you use the ODS OUTPUT statement to
save the table as an output data set. You can also display this column by modifying the corresponding
template.

### BY Statement

```
BY variables ;
```

You can specify a BY statement in PROC CAUSALTRT 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 in the CAUSALTRT proce-
dure. 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 *Base SAS Procedures Guide.*

### CLASS Statement

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

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

The CLASS statement must precede the MODEL and PSMODEL 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 classification variable. If both the DESCENDING and ORDER= options are specified, PROC CAUSALTRT 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 label 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 variable.

ORDER=DATA | FORMATTED | FREQ | INTERNAL
specifies the sort order for the levels of classification variables. This ordering determines which parameters in the model correspond to each level in the data.

The following table shows how PROC CAUSALTRT interprets values of the ORDER= option.

<table>
<thead>
<tr>
<th>Value of ORDER=</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 values, except for numeric variables that have no explicit format, which are sorted by their unformatted (internal) values. The sort order is machine-dependent.</td>
</tr>
<tr>
<td>FREQ</td>
<td>Descending frequency count. Levels that have more observations come earlier in the order.</td>
</tr>
<tr>
<td>INTERNAL</td>
<td>Unformatted value. The sort order is machine-dependent.</td>
</tr>
</tbody>
</table>

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

REF='level' | FIRST | LAST
specifies a level of the classification 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 reference.

LAST designates the last ordered level as reference.

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.

---

**FREQ Statement**

FREQ variable;

The variable in the FREQ statement identifies a variable in the input data set that contains the frequency of occurrence of each observation. PROC CAUSALTRT treats each observation as if it appeared \( n \) times, where \( n \) is the value of the FREQ variable for the observation. If the frequency value is not an integer, it is truncated to an integer. If it is less than 1 or missing, the observation is not used.

---

**MODEL Statement**

MODEL outcome < (variable-options)> <= effects> < / model-options> ;

For all the estimation methods that PROC CAUSALTRT implements, you must specify the outcome variable in the MODEL statement. For more information about the estimation methods implemented in PROC CAUSALTRT and the model specifications that are required, see the section “Outline of Estimation Method Requirements” on page 2367.

In addition to specifying the outcome variable, you can also specify the following:

**effects**

specify one or more explanatory variables or a combination of variables that are used to fit regression models for the outcome variable within each treatment condition. Explanatory variables that represent nominal (classification) data must be declared in the CLASS statement. For more information about specifying effects, see the section “Specification of Effects” on page 4020 in Chapter 50, “The GLM Procedure.”

**variable-options**

specify one or more of the following options within parentheses for a binary outcome variable immediately after the variable:

**DESCENDING**

DESC

reverses the order of the outcome categories. If both the DESCENDING and ORDER= options are specified, PROC CAUSALTRT orders the outcome categories according to the ORDER= option and then reverses that order.

**EVENT='category' | FIRST | LAST**

specifies the event category for the binary outcome model. PROC CAUSALTRT models the probability of the event category. You can specify one of the following:
'category' specifies the value (formatted if a format is applied) of the event category in quotation marks.

**FIRST** designates the first-ordered category as the event.

**LAST** designates the last-ordered category as the event.

By default, EVENT=FIRST.

One of the most common sets of outcome levels is \{0,1\}, where 1 represents the event for which the probability is to be modeled. Consider the following example, where \( Y \) takes the values 1 and 0 for event and nonevent, respectively, and \( X \) is the explanatory variable. To specify the value 1 as the event category, use the following MODEL statement:

```plaintext
model Y(event='1') = X;
```

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

specifies the sort order for the levels of the outcome variable. The following table displays the available ORDER= options.

<table>
<thead>
<tr>
<th>ORDER=</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, ORDER=FORMATTED.

For more information about sort order, see the chapter on the SORT procedure in the *Base SAS Procedures Guide* and the discussion of BY-group processing in *SAS Language Reference: Concepts*.

**REFERENCE='category' | FIRST | LAST**

specifies the reference category for the binary outcome model. Specifying one outcome category as the reference is the same as specifying the other outcome category as the event category. You can specify one of the following:

- **'category'** specifies the value (formatted if a format is applied) of the reference category in quotation marks.
- **FIRST** designates the first-ordered category as the reference.
- **LAST** designates the last-ordered category as the reference.

By default, REF=LAST.
specify additional options for the outcome model after a slash (/). These model-options are summarized in Table 36.4 and described in detail after the table.

**Table 36.4**  model-options in the MODEL Statement

<table>
<thead>
<tr>
<th>model-option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA=</td>
<td>Specifies the level for confidence limits</td>
</tr>
<tr>
<td>DIST=</td>
<td>Specifies the probability distribution</td>
</tr>
<tr>
<td>LINK=</td>
<td>Specifies the link function</td>
</tr>
<tr>
<td>NOSCALE</td>
<td>Holds the scale parameter fixed</td>
</tr>
<tr>
<td>SCALE=</td>
<td>Specifies the value used for the scale</td>
</tr>
</tbody>
</table>

**ALPHA=number**  
specifies a number to use as the $\alpha$ level for 100(1 - $\alpha$)% confidence limits that the MODEL statement computes. The value of number must be between 0 and 1. The default value of number is the set by the ALPHA= option in the PROC CAUSALTRT statement.

**DIST=keyword**  
**DISTRIBUTION=keyword**  
specifies the built-in probability distribution to use in the model. If you specify the DIST= option and you omit the LINK= option, a default link function is chosen as displayed in Table 36.5. If you specify neither the DIST= option nor the LINK= option, then the CAUSALTRT procedure defaults to the binomial 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 CAUSALTRT procedure defaults to the normal distribution with the identity link function.

**Table 36.5**  Distributions and Default Link Functions

<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>GAMMA</td>
<td>GAM</td>
<td>Gamma</td>
</tr>
<tr>
<td>IGAUSSIAN</td>
<td>IG</td>
<td>Inverse Gaussian</td>
</tr>
<tr>
<td>NORMAL</td>
<td>NOR</td>
<td>Normal</td>
</tr>
<tr>
<td>POISSON</td>
<td>POI</td>
<td>Poisson</td>
</tr>
</tbody>
</table>

Responses for the inverse Gaussian and gamma distributions must be strictly positive. For the Poisson distribution, 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 causal effect, even if the estimation method does not fit an outcome model.

**LINK=keyword**  
specifies the link function in the model. You can specify the keywords shown in Table 36.6.
Table 36.6  Built-In Link Functions of the CAUSALTRT Procedure

<table>
<thead>
<tr>
<th>LINK=</th>
<th>Link Function</th>
<th>g(μ) = η =</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOGLOG</td>
<td>Complementary log-log</td>
<td>log(-log(1 - μ))</td>
</tr>
<tr>
<td>CLL</td>
<td>Identity</td>
<td>μ</td>
</tr>
<tr>
<td>INVERSE</td>
<td>Reciprocal</td>
<td>1/μ</td>
</tr>
<tr>
<td>RECIPROCAL</td>
<td>Log</td>
<td>log(μ)</td>
</tr>
<tr>
<td>LOGIT</td>
<td>Logit</td>
<td>log(μ/(1 - μ))</td>
</tr>
<tr>
<td>POWERMINUS2</td>
<td>Power with exponent -2</td>
<td>1/μ²</td>
</tr>
<tr>
<td>PROBIT</td>
<td>Probit</td>
<td>Φ⁻¹(μ)</td>
</tr>
</tbody>
</table>

For the probit link, Φ⁻¹(·) denotes the quantile function of the standard normal distribution. By default, the link function is chosen as shown in Table 36.5.

**NOSCALE**

holds the scale parameter fixed. If you omit the SCALE= option, the scale parameter is fixed at the value 1. If you do not specify the NOSCALE option, the scale parameter is estimated by maximum likelihood for the normal, inverse Gaussian, and gamma distributions.

**SCALE=number**

specifies the value used for the scale parameter when the NOSCALE option is specified. For the binomial and Poisson distributions, which have no free scale parameter, you can use this option to specify an overdispersed model. If the NOSCALE option is not specified, then number is used as an initial estimate of the scale parameter.

**OUTPUT Statement**

```
OUTPUT <OUT=SAS-data-set> <keyword=name . . . keyword=name> ;
```

The OUTPUT statement creates a new SAS data set that contains all the variables in the input data set and optionally contains predicted values that are used to estimate the treatment effect. If the estimation of the causal effect requires that a regression model for the outcome variable be fit, then you can request that the predicted potential outcomes be included in the output data set. You can request that the predicted propensity scores and inverse probability weights be included if the estimation of the causal effect fits a model for the treatment assignment.

The output data set contains estimated potential outcomes and propensity scores for all observations in which the explanatory variables for the model are not missing, regardless of whether the outcome is missing.

This behavior enables you to predict potential outcomes and propensity scores of “new” observations without affecting the original model fit. These new observations must have nonmissing explanatory variable values and treatment assignment, but their outcome values are missing in the input data set. This way, these new observations cannot be used for model fitting but their predicted values are computed in the OUT= data set.
You can specify the following options:

\textbf{OUT= SAS-data-set}

names the output data set. If you omit this option, an output data set is created and given a default name that uses the DATA convention.

\textit{keyword= name}

specifies a statistic to be included in the output data set and assigns the specified \textit{name} to the new variable that contains the statistic. Specify one or more of the following \textit{keywords} and \textit{names} (connect each \textit{keyword} and \textit{name} by an equal sign):

\begin{itemize}
  \item \textbf{IPW= name} requests the predicted inverse probability weight.
  \item \textbf{POCNT= name}
  \item \textbf{PREDCNT= name} requests the predicted potential outcome for the control condition.
  \item \textbf{POTRT= name}
  \item \textbf{PREDTRT= name} requests the predicted potential outcome for the treatment condition.
  \item \textbf{PSCORE= name}
  \item \textbf{PS= name} requests the predicted propensity score.
\end{itemize}

If you do not specify any \textit{keyword= name} options, the output data set contains only the original variables.

\section*{PSMODEL Statement}

\texttt{PSMODEL treatment < (variable-options) >= effects < / psmodel-options >> ;}

For all the estimation methods that PROC CAUSALTRT implements, you must specify a \textit{treatment} variable in a PSMODEL statement. The regression model that is fit by the PSMODEL statement is called the propensity score model. The propensity score refers to the probability of receiving treatment conditional on the model \textit{effects}. For more information about the estimation methods that PROC CAUSALTRT implements, see the section “Estimating the Average Treatment Effect (ATE)” on page 2397.

In addition to specifying the \textit{treatment} variable, you can also specify the following:

\textit{effects}

specify one or more explanatory variables or a combination of variables that are used to fit a regression model for the \textit{treatment} variable. Explanatory variables that represent nominal (classification) data must be declared in the CLASS statement. For more information about specifying \textit{effects}, see the section “Specification of Effects” on page 4020 in Chapter 50, “The GLM Procedure.”
specify one or more of the following options within parentheses for after the \textit{treatment} variable:

\begin{description}
\item[\textbf{DESCENDING}]\textbf{DESC} reverses the order of the treatment categories. If both the \textbf{DESCENDING} and \textbf{ORDER=} options are specified, \texttt{PROC CAUSALTRT} orders the treatment categories according to the \textbf{ORDER=} option and then reverses that order.
\item[\textbf{EVENT}]=\texttt{\textquotesingle category\textquotesingle} | \textbf{FIRST} | \textbf{LAST} specifies the event category for the binary treatment model. \texttt{PROC CAUSALTRT} models the probability of the event category. You can specify one of the following:
\begin{itemize}
\item \texttt{\textquotesingle category\textquotesingle} specifies the value (formatted if a format is applied) of the event category in quotation marks.
\item \textbf{FIRST} designates the first-ordered category as the event.
\item \textbf{LAST} designates the last-ordered category as the event.
\end{itemize}
By default, \texttt{EVENT=}\texttt{FIRST}.
\end{description}

One of the most common sets of treatment levels is \{0,1\}, where 1 represents the event for which the probability is to be modeled. Consider the following example, where \texttt{T} takes the values 1 and 0 for event and nonevent, respectively, and \texttt{X} is the explanatory variable. To specify the value 1 as the event category, use the following \texttt{PSMODEL} statement:

\begin{verbatim}
psmodel T(event='1') = X;
\end{verbatim}

\begin{description}
\item[\textbf{ORDER=}]=\texttt{DATA} | \texttt{FORMATTED} | \texttt{FREQ} | \texttt{INTERNAL} specifies the sort order for the levels of the treatment variable. The following table displays the available \textbf{ORDER=} options.

<table>
<thead>
<tr>
<th>ORDER=</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</td>
</tr>
<tr>
<td></td>
<td>have no explicit format, which are sorted by their unformatted</td>
</tr>
<tr>
<td></td>
<td>(internal) value. The sort order is machine-dependent.</td>
</tr>
<tr>
<td>FREQ</td>
<td>Descending frequency count. Levels that have the most</td>
</tr>
<tr>
<td></td>
<td>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>
\end{description}

By default, \texttt{ORDER=}\texttt{FORMATTED}.

For more information about sort order, see the chapter on the \texttt{SORT} procedure in the \textit{Base SAS Procedures Guide} and the discussion of \texttt{BY}-group processing in \textit{SAS Language Reference: Concepts}. 
REFERENCE='category' | FIRST | LAST
REF='category' | FIRST | LAST

specifies the reference category for the binary treatment model. Specifying one treatment category as the reference is the same as specifying the other treatment category as the event category. You can specify one of the following:

'category' specifies the value (formatted if a format is applied) of the reference category in quotation marks.
FIRST designates the first-ordered category as the reference.
LAST designates the last-ordered category as the reference.

By default, REF=LAST.

psmodel-options

specify additional options for the propensity score model after a slash (/). These *psmodel-options* are summarized in Table 36.7 and described in detail after the table. These options are applied only if the estimation method you request requires fitting a propensity score model.

**Table 36.7 psmodel-options in the PSMODEL Statement**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA=number</td>
<td>Specifies the level for confidence limits</td>
</tr>
<tr>
<td>PLOTS=</td>
<td>Produces graphics for the propensity scores and weights</td>
</tr>
<tr>
<td>WGTFLAG=</td>
<td>Specifies a value to be used to flag large weights</td>
</tr>
</tbody>
</table>

**ALPHA=number**

specifies a number to use as the $\alpha$ level for $100(1-\alpha)\%$ confidence limits that the PSMODEL statement computes. The value of number must be between 0 and 1. The default value of number is the set by the ALPHA= option in the PROC CAUSALTRT statement.

**PLOTS < (global-plot-options) >=plot-request< (options) >**

produces ODS graphics by using the predicted propensity scores. No graphics are produced if you specify the PLOTS=NONE option in the PROC CAUSALTRT statement or if the estimation method used does not require fitting a propensity score model.

ODS Graphics must be enabled before plots can be requested. For example:

```plaintext
ods graphics on;

proc causaltrt method=ipwr;
   model y;
      psmodel trt = x1 x2 /plots=all;
run;

ods graphics off;
```
For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 623 in Chapter 21, “Statistical Graphics Using ODS.”

You can specify the following plot-requests:

**ALL**
produces all possible plots.

**LOGITPSCORE**
**LPS**
produces overlaid density plots for the logit of the propensity score within each treatment condition.

**NONE**
suppresses all plots that the PSMODEL statement could produce. This plot-request is used to suppress output if the PLOTS=ALL option is specified in the PROC CAUSALTRT statement.

**OUTBYPSCORE**
**OUTBYS**
produces a scatter plot of the outcome variable by the propensity score for nonbinary outcomes. If the outcome is binary, box plots of the propensity scores within treatment conditions are produced for each outcome level. The whisker lengths for each box plot are determined by the maximum and minimum of the propensity scores.

**OUTBYWEIGHT**
**OUTBYWGT**
produces a scatter plot of the outcome variable by weight for nonbinary outcomes. If the outcome is binary, box plots of the weights within treatment conditions are produced for each outcome level. The whisker lengths for each box plot are determined by the maximum and minimum of the weights.

**PSCLOUD**
produces a point cloud of the propensity scores by jittering within the control and treatment conditions.

**PSCOV DEN**< pscovden-options >
produces density plots for the covariates or continuous effects that are specified in the PSMODEL statement. Each plot displays the density of effects for the treatment and control condition. Two plots are produced for each effect: one plot displays unweighted densities, and the other plot displays densities that are weighted by inverse probability weights. By default, plots are produced for all continuous effects that are specified in the PSMODEL statement, and the plots are collected in panels. You can specify the following pscovden-options:

**EFFECTS(effects)**
specifies the effects for which density plots are produced. Plots are produced only for continuous effects that are specified in the PSMODEL statement. The effects consist of an explanatory variable or combination of variables. The syntax for specifying effects in this option is the same as the syntax for specifying effects in the MODEL and PSMODEL statements.
UNPACK
suppresses paneling. By default, multiple plots can appear in the same output panel. You can use this option to display each plot separately.

UNPACKEFFECTS
creates a separate panel for each effect. Each panel has a single row and two columns that contain the unweighted and weighted density plots. This option is ignored if you specify the UNPACK option in the PROC CAUSALTRT or PSMODEL statements.

PSDIST
produces a box plot of the propensity scores for each treatment condition. The whisker lengths for each box plot are determined by the maximum and minimum of the propensity scores.

WEIGHTCLOUD
WCloud
produces a point cloud of the weights by jittering within the control and treatment conditions.

WEIGHTDIST
WDist
produces a box plot of the weights for each treatment condition. The whisker lengths for each box plot are determined by the maximum and minimum of the weights.

You can specify the following global-plot-options:

ONLY
suppresses plots that are not specified as a plot-request. This option applies only if the PLOTS= option is specified in the PROC CAUSALTRT statement.

UNPACK
UNPACKPANEL
suppresses paneling. By default, multiple plots can appear in the same output panel. You can specify UNPACK to display each plot separately.

WGTFLAG=number
specifies a number to use to flag large weights. The value of number must be greater than or equal to 1. If any observations have weights greater than number, a note is added to the SAS log. By default, WGTFLAG=50.

Details: CAUSALTRT Procedure

Causal Effects: Definitions, Assumptions, and Identification

The CAUSALTRT procedure estimates the causal effect of a binary treatment, T, on a continuous or discrete outcome, Y. This section defines causal effects and discusses the assumptions and conditions that enable researchers to establish causal interpretations.
Although the causal effects that PROC CAUSALTRT estimates are called treatment effects, the variable $T$ might represent an intervention or exposure, depending on the problem nature. For example, consider that $T$ is an intervention such as receiving subsidized access to a fitness center, and $Y$ is measure of physical health. The researcher would like to assess the treatment or causal effect of the subsidized fitness program $T$ on the physical health $Y$. But what exactly is the causal effect? This section uses Neyman-Rubin’s potential outcomes framework (Rubin 1980, 1990) to define causal effects and to discuss their assumptions. Next, it explains the conditions that enable researchers to identify and establish causal effects.

The first important concept for defining a causal effect is potential outcomes. Potential outcomes describe an idealized source of data where you can observe a subject’s response for all possible treatment assignments. For example, if you are estimating the causal effect of a binary variable for which $T = 0$ indicates assignment to the control and $T = 1$ assignment to treatment, then each subject has two potential outcomes $Y(0)$ and $Y(1)$. The subject-level (also called unit-level) treatment effect is defined as the difference in potential outcomes $Y(1) - Y(0)$.

To formally define the potential outcomes as unique and intrinsic values that are associated with a subject, Rubin (1980) states the stable unit treatment value assumption (SUTVA). See also Imbens and Rubin (2015). Two components of the SUTVA are the following:

- No interference. A subject’s potential outcomes are not affected by the treatment assignments of the other subjects.
- No hidden variations of treatments. Subjects must receive the same form of treatment at each treatment level.

With a more refined statement, VanderWeele (2009) proposed a less restrictive version to the second component of SUTVA. Essentially, VanderWeele’s version requires that the potential outcomes remain unchanged within certain degree of treatment variation for each treatment level. This version is also known as the treatment variation irrelevance assumption. See also Cole and Frangakis (2009).

Furthermore, Rubin (2005) points out an implicit but important assumption for constructing the potential outcome framework: the potential outcomes are not affected by the treatment assignment mechanism that the researchers use to learn about them. Whether in an observational study or a randomized experiment, potential outcomes remain unique to the subjects given the well-defined causal problem and SUTVA.

To summarize, the potential outcome framework with SUTVA ensures that the causal treatment effect is well defined. However, to actually attempt to estimate causal treatment effects from the data, the consistency assumption that relates an observed outcome to the potential outcome is needed:

$$ Y = TY(1) + (1 - T)Y(0) $$

This equation states that the observed response $Y$ is equal to the potential outcome with a treatment level that matches the actual assigned treatment level. In this setting, if $j$ is an unobserved treatment condition, the unobserved potential outcome $Y(j)$ is defined using counterfactual conditionals as the outcome that would occur if, contrary to the observed assignment, the subject received treatment $T = j$. Because each subject can take part in only one treatment condition, at least half of the potential outcomes would not be observed in the data collection process. Therefore, the unit-level causal effect, as defined by $Y(1) - Y(0)$, is seldom the primary interest in research. Instead, average treatment effects in the population are the more common estimands.

The CAUSALTRT procedure estimates two types of treatment effects:
The average treatment effect (ATE) for the entire population is given by
\[
\text{ATE} = \mu_1 - \mu_0 = E[Y(1)] - E[Y(0)]
\]
where \( \mu_1 = E[Y(1)] \) and \( \mu_0 = E[Y(0)] \) are potential outcome means (POMs) for the treatment and control conditions, respectively.

The average treatment effect for the treated (ATT or ATET) is the average causal effect among only those individuals that receive treatment and is given by
\[
\text{ATT} = \mu_{1|T=1} - \mu_{0|T=1} = E[Y(1)|T = 1] - E[Y(0)|T = 1]
\]
where \( \mu_{1|T=1} = E[Y(1)|T = 1] \) and \( \mu_{0|T=1} = E[Y(0)|T = 1] \) are potential outcome means for the treatment and control conditions, respectively, conditional on receiving treatment.

It is important to notice that the preceding discussion of potential outcome framework, the associated assumptions, and the definitions of causal treatment effects apply to both experimental and observational data (Rubin 1990). But for estimating causal treatment effects, statistical methodology for experimental and observational data would be differentiated.

The differentiation is caused by the following important condition (or assumption) for the identification of causal treatment effect: given the pretreatment covariates \( X \), the potential outcomes are independent of treatment assignment.

This condition states that if all the covariates \( X \) have been included in the model appropriately (that is, there are no unmeasured covariates), the treatment effect of \( T \) on \( Y \) conditional on \( X \) would have a causal interpretation. Consequently, the causal treatment effects can be estimated as some functions of the observed conditional means.

Unfortunately, this condition cannot be 100% established in most applications; hence, this condition is usually referred to as an assumption in causal analysis literature. This assumption is called strong ignorability of treatment assignment or conditional exchangeability. This assumption sheds light on the different statistical methodology between randomized experiments and observational studies.

In experiments where randomization is used to assign subjects to treatment conditions, you can safely assume that the treatment assignment and potential outcomes are independent by design. Hence, for randomized experiments, traditional statistical methods such as \( t \) tests and ANOVA can be used without worrying about the confounding that can be caused by covariates \( X \).

However, in observational studies, subjects “select” the treatment conditions based on their pretreatment characteristics (covariates), \( X \), which could also be associated with the outcome variable. The treatment and outcome association would be confounded by the covariates. In this case, identification of causal effects requires that the strong ignorability or conditional exchangeability assumption be satisfied; that is, only when the estimation of the causal effects can successfully take into account all the confounding that is caused by the covariates \( X \).

The CAUSALTRT procedure implements several statistical techniques that enable you to estimate treatment effects when you specify confounding covariates \( X \). PROC CAUSALTRT uses the covariates \( X \) to fit generalized linear models for the treatment \( T \), outcome \( Y \), or both. Predicted values from these models are then incorporated into the estimation of the causal effects. In particular, you can estimate the ATE by the following methods:
• inverse probability weighting, where weights are predicted from the model for the treatment assignment $T$ (also called the propensity score model)

• regression adjustment, where the model of the outcome $Y$ is used to predict potential outcomes for each treatment assignment

• doubly robust estimation, which combines both inverse probability weighting and regression adjustment methods to estimate the causal effect; doubly robust estimation methods provide unbiased estimates even if one of the models is misspecified

Whenever your modeling involves a model for the treatment assignment $T$ or propensity score (that is, when you use either the inverse probability weighting or doubly robust estimation), an additional assumption about positivity is needed: each subject should have a nonzero probability of receiving any treatment condition. This assumption is expressed as follows:

$$0 < \Pr(T = 1 \mid X = x) < 1$$

For more information about estimation of the ATE, see the section “Estimating the Average Treatment Effect (ATE)” on page 2397. For more information about the models that are required for the estimation methods that PROC CAUSALTRT implements and how default estimation methods are determined, see the section “Outline of Estimation Method Requirements” on page 2367. You can estimate the ATT by using either inverse probability weights or regression adjustment. For more information about estimation of the ATT, see the section “Estimating the Average Treatment Effect for the Treated (ATT)” on page 2402.

### Estimating the Average Treatment Effect (ATE)

The CAUSALTRT procedure can estimate the average treatment effect (ATE) by using inverse probability weighting, regression adjustment, and doubly robust methods (doubly robust methods are combinations of the first two methods). The following sections describe each of these estimation methods.

#### Inverse Probability Weighting

The CAUSALTRT procedure estimates the potential outcome means and ATE by using the three inverse probability weighting methods that are described in Lunceford and Davidian (2004). All three methods compute weights using estimates for the conditional probability of receiving treatment,

$$e(x) = \Pr(T = 1 \mid X = x)$$

where $x$ is a vector of the observed covariates. The conditional probability $e(x)$ is called the propensity score, and the model that is used to estimate $e(x)$ is called the propensity score model. The inverse probability weight for an observation is equal to

$$\frac{1}{\Pr(T = t \mid x)} = \frac{t}{e(x)} + \frac{1 - t}{1 - e(x)}$$

In addition to the stable unit treatment value assumption (SUTVA), two more assumptions on the propensity scores are required for the use of inverse probability weighting methods. The positivity assumption, that $0 < e(x) < 1$, ensures that each observation has a nonzero probability of receiving each treatment condition.
The assumption of no unmeasured confounders, also known as strong ignorability, requires that the potential outcomes $Y(0)$ and $Y(1)$ be independent of the treatment variable $T$ conditional on the covariates $x$. In practice, this assumption means that the set of covariates $x$ should have included all the important confounders in the propensity score model. If these conditions are satisfied and the propensity score model is correctly specified, then it follows that

$$E \left[ \frac{TY}{e(x)} \right] - E \left[ \frac{(1-T)Y}{1-e(x)} \right] = E \left[ \frac{Y(1)}{e(x)} E[T \mid x] \right] - E \left[ \frac{Y(0)}{1-e(x)} E[(1-T) \mid x] \right] = E[Y(1)] - E[Y(0)].$$

The CAUSALTRT procedure uses the effects that are specified in the PSMODEL statement to fit a logistic regression to the propensity score model. The inverse probability weights are then computed by taking the inverse of the estimated propensity scores.

Suppose you have observations $(y_i, t_i, x_i)$, for $i = 1, \ldots, n$. The parameter estimates for the propensity score model is given by $\hat{\mathbf{\beta}}_{ps}$, and the predicted values for the propensity score are given by

$$\hat{e}_i = \hat{e}(x_i) = \frac{\exp \left( x_i' \hat{\mathbf{\beta}}_{ps} \right)}{1 + \exp \left( x_i' \hat{\mathbf{\beta}}_{ps} \right)}$$

The three weighting methods that PROC CAUSALTRT implements obtain unbiased estimates of the potential outcome means by solving the equations

$$S_{ipw}(\mu) = \sum_{i=1}^{n} S_{ipw,i} = 0$$

for $\mu = (\mu_0, \mu_1)$, where

$$S_{ipw,i} = \begin{bmatrix} (1-t_i)(y_i-\mu_0) - \eta_0 \left( \frac{t_i-\hat{e}_i}{1-\hat{e}_i} \right) \\ t_i(y_i-\mu_1) - \eta_1 \left( \frac{t_i-\hat{e}_i}{\hat{e}_i} \right) \end{bmatrix}$$

These estimation methods differ in the values of $(\eta_0, \eta_1)$. The choices of $(\eta_0, \eta_1)$ and the corresponding potential outcome mean estimates for each method are as follows:

- **IPW:**
  $$(\eta_0, \eta_1) = (\mu_0, \mu_1)$$
  $$\hat{\mu}_{0}^{ipw} = n^{-1} \sum_{i=1}^{n} \frac{(1-t_i)y_i}{1-\hat{e}_i}$$
  $$\hat{\mu}_{1}^{ipw} = n^{-1} \sum_{i=1}^{n} \frac{t_iy_i}{\hat{e}_i}$$

- **IPWR:**
  $$(\eta_0, \eta_1) = (0, 0)$$
Estimating the Average Treatment Effect (ATE)

\[
\hat{\mu}_0 = \left[ \sum_{i=1}^{n} \frac{(1 - t_i)}{1 - \hat{e}_i} \right]^{-1} \sum_{i=1}^{n} \frac{(1 - t_i)y_i}{1 - \hat{e}_i}
\]

\[
\hat{\mu}_1 = \left[ \sum_{i=1}^{n} \frac{t_i}{\hat{e}_i} \right]^{-1} \sum_{i=1}^{n} \frac{t_iy_i}{\hat{e}_i}
\]

- **IPWS:**

\[
(\eta_0, \eta_1) = \left( -\frac{E[(1 - T)(Y - \mu_0)/(1 - e(x))^2]}{E[(T - e(x))^2/(1 - e(x))^2]} - \frac{E[T(Y - \mu_1)/e(x)^2]}{E[(T - e(x))^2/e(x)^2]} \right)
\]

\[
\hat{\mu}_0^\text{ipws} = - \left[ \sum_{i=1}^{n} \frac{(1 - t_i)}{1 - \hat{e}_i} \left( 1 - \frac{C_0}{1 - \hat{e}_i} \right) \right]^{-1} \sum_{i=1}^{n} \frac{(1 - t_i)y_i}{1 - \hat{e}_i} \left( 1 - \frac{C_0}{1 - \hat{e}_i} \right)
\]

\[
\hat{\mu}_1^\text{ipws} = \left[ \sum_{i=1}^{n} \frac{t_i}{\hat{e}_i} \left( 1 - \frac{C_1}{\hat{e}_i} \right) \right]^{-1} \sum_{i=1}^{n} \frac{t_iy_i}{\hat{e}_i} \left( 1 - \frac{C_1}{\hat{e}_i} \right)
\]

where

\[
C_0 = \left[ -\sum_{j=1}^{n} \frac{t_j - \hat{e}_j}{1 - \hat{e}_j} \left[ \sum_{j=1}^{n} \frac{t_j - \hat{e}_j}{1 - \hat{e}_j} \right]^2 \right]^{-1}
\]

\[
C_1 = \left[ \sum_{j=1}^{n} \frac{t_j - \hat{e}_j}{\hat{e}_j} \left[ \sum_{j=1}^{n} \frac{t_j - \hat{e}_j}{\hat{e}_j} \right]^2 \right]^{-1}
\]

If the values of \((\eta_0, \eta_1)\) that motivate the IPWS method were known constants, they would produce estimates for the potential outcome means that have the smallest asymptotic variance among the class of estimators that is defined by \(S_{ipw}(\mu)\). The IPWS method estimates \((\hat{\eta}_0, \hat{\eta}_1)\) by solving the estimating equations

\[
S(\eta) = \sum_{i=1}^{n} S_{\eta,i} = 0
\]

for \(\eta = (\eta_0, \eta_1)\), where:

\[
S_{\eta,i} = \left[ \frac{(1-t_i)(y_i-\mu_0)}{(1-\hat{e}_i)^2} + \eta_0 \left( \frac{t_i-\hat{e}_i}{1-\hat{e}_i} \right)^2 \right]
\]

\[
+ \eta_1 \left( \frac{t_i-\hat{e}_i}{\hat{e}_i} \right)^2
\]

Although you do not need to specify an outcome model when you use the inverse probability weighting methods, you still need to specify the outcome variable in the MODEL statement. The IPW, IPWR, or IPWS estimation method can be requested using the METHOD= option in the PROC CAUSALTRT statement. For more information about which model specifications are required for different estimation methods and how default estimation methods are determined, see the section “Outline of Estimation Method Requirements” on page 2367.
The robust sandwich covariance estimate for \( \hat{\mu} \) is computed by stacking the equations \( S_{\text{ipw}}(\mu) \) and the score function for the logistic regression model that is used to obtain \( \hat{\beta}_{\text{ps}} \). By stacking the equations, PROC CAUSALTRT adjusts the covariance matrix for \( \hat{\mu} \) to account for the prediction of \( \hat{\beta}_{\text{ps}} \) and the propensity scores \( \hat{e}_i \). For more information about using the robust sandwich covariance estimate to compute standard errors, see the section “Asymptotic Formulas” on page 2404.

The finite sample properties of \( \hat{\mu} \) and its covariance matrix can be affected by observations that have large weights. By default, a note is added to the SAS log if any observations have weights greater than 50. You can specify the value that is used to flag large weights in the WGTFLAG= option in the PSMODEL statement. You can also inspect the values of the predicted weights by requesting graphics in the PLOTS= option in the PROC CAUSALTRT or PSMODEL statements.

As described in Rosenbaum and Rubin (1983), the propensity score is also a balancing score. To assess the balance that is produced by a propensity score model, you can request the display of weighted and unweighted standardized mean differences and variance ratios for propensity score model effects in the COVDIFFPS option in the PROC CAUSALTRT statement. You can also request weighted and unweighted densities for propensity score model effects by specifying PSCOVDEN in the PLOTS= option in the PROC CAUSALTRT or PSMODEL statements. To inspect the balance of a propensity score model without showing the causal effects, specify the NOEFFECT option in the PROC CAUSALTRT statement.

Based on the balancing score property, you can use the propensity scores to create matched samples or strata for the data. For more information about applying the propensity score in matching and stratification see Chapter 98, “The PSMATCH Procedure.”

**Regression Adjustment**

Regression adjustment estimates the ATE by using predicted values for the potential outcomes. The CAUSALTRT procedure predicts the potential outcomes from generalized linear models that are fit to the data by maximum likelihood estimation. For more information about generalized linear models, see the section “Generalized Linear Models Theory” on page 3491 in Chapter 48, “The GENMOD Procedure.”

The effects that you specify in the MODEL statement are used to fit models for the outcome within each treatment condition. The predicted potential outcomes are then given by

\[
\hat{y}_{i0} = g^{-1}(x_i' \hat{\beta}_c)
\]

\[
\hat{y}_{i1} = g^{-1}(x_i' \hat{\beta}_t)
\]

where \( g \) is the link function being used and \( \hat{\beta}_c \) and \( \hat{\beta}_t \) are the parameter estimates for the control and treatment outcome models, respectively. The potential outcome mean \( \mu_j \) is then estimated by averaging the predicted \( \hat{y}_{ij} \) for all observations \( i = 1, \ldots, n \), regardless of the observed treatment assignment. Therefore, the regression adjustment estimates for the potential outcome means solve the estimating equations

\[
S_{\text{reg}}(\mu) = \sum_{i=1}^{n} S_{\text{reg},i} = 0
\]

where

\[
S_{\text{reg},i} = \begin{bmatrix} \hat{y}_{i0} - \mu_0 \\ \hat{y}_{i1} - \mu_1 \end{bmatrix}
\]
The regression adjustment estimates for the potential outcome means are equal to

\[
\hat{\mu}_0^{\text{reg}} = n^{-1} \sum_{i=1}^{n} \hat{y}_{i0} \\
\hat{\mu}_1^{\text{reg}} = n^{-1} \sum_{i=1}^{n} \hat{y}_{i1}
\]

If the outcome model is correctly specified and the assumption of no unmeasured confounders is satisfied, then regression adjustment produces unbiased estimates for the potential outcome means. The outcome model effects should also have similar distributions between treatment conditions to avoid extrapolation when the potential outcomes \( \hat{y}_{ij} \) are predicted.

Although you do not need to specify a propensity score model in order to estimate the potential outcome means by the regression adjustment method, you still need to specify the treatment variable in the PSMODEL statement. You can specify METHOD=REGADJ in the PROC CAUSALTRT statement to request the regression adjustment method. For more information about which model specifications are required for different estimation methods and how default estimation methods are determined, see the section “Outline of Estimation Method Requirements” on page 2367.

The robust sandwich covariance estimate for \( \hat{\mu}^{\text{reg}} \) is estimated by stacking the equations \( S_{\text{reg}}(\mu) \) and the score function for the outcome models that are used to estimate \( \hat{\beta}_c \) and \( \hat{\beta}_t \). By stacking the equations, PROC CAUSALTRT adjusts the covariance matrix for \( \hat{\mu}^{\text{reg}} \) to account for the prediction of the potential outcomes \( \hat{y}_{ij} \). For more information about using the robust sandwich covariance estimate to compute standard errors, see the section “Asymptotic Formulas” on page 2404.

**Doubly Robust Estimation**

Doubly robust estimation methods fit models for both the outcome and treatment variables. They combine inverse probability weighting and regression adjustment to estimate the potential outcome means. The methods are said to be doubly robust because they provide unbiased estimates for \( \mu \) even if one of the models is misspecified (Bang and Robins 2005). In this sense, you have two opportunities to obtain unbiased estimation of causal effects by specifying either a correct treatment or a correct outcome model. The CAUSALTRT procedure implements two doubly robust estimation methods: the augmented inverse probability weighting (AIPW) method described in Lunceford and Davidian (2004) and the inverse probability weighted regression adjustment (IPWREG) method described in Wooldridge (2010).

The AIPW estimation method estimates propensity scores \( \hat{e}_i \) and the associated parameter estimates \( \hat{\beta}_{ps} \) by fitting a logistic regression model for treatment assignment, as described in section “Inverse Probability Weighting” on page 2397. It also estimates the potential outcomes \( \hat{y}_{ij} \) by using the maximum likelihood fitting of the generalized linear models, as described in section “Regression Adjustment” on page 2400. Note that the treatment assignment and outcome models do not need to be fit using the same set of model effects.

The AIPW estimates for potential outcome means solve the estimating equations

\[
S_{\text{aipw}}(\mu) = \sum_{i=1}^{n} S_{\text{aipw},i}(\mu) = 0
\]

for \((\mu_0, \mu_1)\), where

\[
S_{\text{aipw},i}(\mu) = \begin{bmatrix}
\frac{(1-t_i) y_i}{1-\hat{e}_i} + \hat{y}_{i0} \left( \frac{t_i - \hat{e}_i}{1 - \hat{e}_i} \right) - \mu_0 \\
\frac{t_i y_i}{\hat{e}_i} - \hat{y}_{i1} \left( \frac{t_i - \hat{e}_i}{\hat{e}_i} \right) - \mu_1 \\
\hat{y}_{i0} - \left( \frac{1-t_i}{1-\hat{e}_i} \right) (\hat{y}_{i0} - y_i) - \mu_0 \\
\hat{y}_{i1} - \left( \frac{t_i}{\hat{e}_i} \right) (\hat{y}_{i1} - y_i) - \mu_1
\end{bmatrix}
\]
The AIPW estimates for the potential outcome means are given by

\[
\hat{\mu}_{0}^{\text{aipw}} = n^{-1} \sum_{i=1}^{n} (1 - t_i) \frac{y_i}{1 - \hat{e}_i} + \hat{y}_i 0 \left( \frac{t_i - \hat{e}_i}{1 - \hat{e}_i} \right)
\]

\[
\hat{\mu}_{1}^{\text{aipw}} = n^{-1} \sum_{i=1}^{n} t_i y_i \left( \frac{t_i - \hat{e}_i}{\hat{e}_i} \right) - \hat{y}_i 1 \left( \frac{t_i - \hat{e}_i}{\hat{e}_i} \right)
\]

If one of the models is misspecified, \( \hat{\mu}^{\text{aipw}} \) remains an unbiased estimate of \( \mu \). However, the efficiency of \( \hat{\mu}^{\text{aipw}} \) and the estimation of its covariance matrix are affected.

The IPWREG estimation method uses the same logistic model for predicting \( \hat{e}_i \), but it fits weighted models for the outcome variable separately for each treatment condition. To fit the outcome models, each observation is weighted by the inverse probability weights as follows:

\[
\frac{1}{\Pr (T = t_i \mid x_i)} = \frac{t_i}{e(x_i)} + \frac{1 - t_i}{1 - e(x_i)}
\]

Let \( \hat{\beta}_w^{\text{c}} \) and \( \hat{\beta}_w^{\text{t}} \) denote the parameter estimates from the weighted regression outcome models for the control and treatment conditions, respectively. The predicted potential outcomes are then given by

\[
\hat{y}_i 0 = g^{-1} (x_i' \hat{\beta}_w^{\text{c}})
\]

\[
\hat{y}_i 1 = g^{-1} (x_i' \hat{\beta}_w^{\text{t}})
\]

where \( g \) is the link function being used. For the IPWREG method, PROC CAUSALTRT enforces the use of the canonical link for the distribution of the outcome variable. You cannot override the canonical link by the LINK= option. For information about the canonical link for each distribution, see Table 36.5.

The IPWREG estimates for the potential outcome means are given by

\[
\hat{\mu}_{0}^{\text{wreg}} = n^{-1} \sum_{i=1}^{n} \hat{y}_i 0
\]

\[
\hat{\mu}_{1}^{\text{wreg}} = n^{-1} \sum_{i=1}^{n} \hat{y}_i 1
\]

These estimates solve the following estimating equations:

\[
S_{\text{reg}}^w (\mu) = \sum_{i=1}^{n} S_{\text{reg},i}^w = \sum_{i=1}^{n} \left[ \frac{\hat{y}_i 0 - \hat{\mu}_0}{\hat{y}_i 0 - \hat{\mu}_1} \right] = 0
\]

**Estimating the Average Treatment Effect for the Treated (ATT)**

In some research settings or program evaluation studies, instead of the average treatment effect (ATE), researchers or policy makers might be more interested in the causal treatment effects only for those who choose to participate in the treatment condition. Hence, the average treatment effect for the treated (ATT or ATET) becomes the focus.

The CAUSALTRT procedure enables you to estimate the ATT and the corresponding conditional potential outcome means by using either of the following methods:
• Inverse probability weighting with ratio adjustment (IPWR). To estimate the ATT, the inverse probability weights that are described in the section “Inverse Probability Weighting” on page 2397 are multiplied by the predicted propensity scores. The ATT weights are therefore given by

\[
e(x_i) \Pr (T = t_i | x_i) = t_i + \frac{(1 - t_i) e(x_i)}{1 - e(x_i)}
\]

The estimating equations that are solved by the IPWR estimates for the conditional potential outcome means \(\mu_{T=1}\) are

\[
S_{\text{ipwr}}(\mu_{T=1}) = \sum_{i=1}^{n} S_{\text{ipwr},i|T=1} = 0
\]

where

\[
S_{\text{ipwr},i|T=1} = \begin{bmatrix}
(1 - t_i)(y_i - \mu_{0|T=1}) \\
t_i(y_i - \mu_{1|T=1})
\end{bmatrix}
\]

The IPWR estimates for the conditional potential outcome means \(\mu_{T=1}\) are given by

\[
\hat{\mu}_{0|T=1}^\text{ipwr} = \left[ \sum_{i=1}^{n} (1 - t_i) \frac{\hat{e}_i}{1 - \hat{e}_i} \right]^{-1} \sum_{i=1}^{n} (1 - t_i) y_i \frac{\hat{e}_i}{1 - \hat{e}_i}
\]

\[
\hat{\mu}_{1|T=1}^\text{ipwr} = \left[ \sum_{i=1}^{n} t_i \right]^{-1} \sum_{i=1}^{n} t_i y_i
\]

If the propensity score model is correctly specified and the stable unit treatment value assumption (SUTVA), positivity, and no unmeasured confounders assumptions are satisfied, then the predicted conditional means are unbiased estimates for \(\mu_j|T=1\).

• Regression adjustment (REGADJ). For this method, PROC CAUSALTRT obtains predicted potential outcomes \(\hat{y}_{ij}\) from the outcome models that are fitted separately for each treatment condition. This approach is described in section “Regression Adjustment” on page 2400. To estimate \(\mu_{T=1}\), the predicted values are averaged only for individuals who received treatment. The REGADJ estimates for the conditional potential outcome means \(\mu_{T=1}\) are given by

\[
\hat{\mu}_{0|T=1}^\text{reg} = \left[ \sum_{i=1}^{n} t_i \right]^{-1} \sum_{i=1}^{n} t_i \hat{y}_{i0}
\]

\[
\hat{\mu}_{1|T=1}^\text{reg} = \left[ \sum_{i=1}^{n} t_i \right]^{-1} \sum_{i=1}^{n} t_i \hat{y}_{i1}
\]

The estimates therefore solve the estimating equations

\[
S_{\text{reg}}(\mu_{T=1}) = \sum_{i=1}^{n} S_{\text{reg},i|T=1} = 0
\]

where

\[
S_{\text{reg},i|T=1} = \begin{bmatrix}
t_i \left( \hat{y}_{i0} - \mu_{0|T=1} \right) \\
t_i \left( \hat{y}_{i1} - \mu_{1|T=1} \right)
\end{bmatrix}
\]

To request the estimation of the conditional potential outcome means and ATT, specify the ATT option in the PROC CAUSALTRT statement.
Comparing Weights for Estimating the ATE and ATT

This section describes and distinguishes the weights that are used in different estimation methods and analysis situations.

When the average treatment effect (ATE) is estimated, the inverse probability weight for an observation is equal to

\[
\frac{1}{\Pr(T = t_i | x_i)} = \frac{t_i}{e(x_i)} + \frac{1 - t_i}{1 - e(x_i)}
\]

These are the weights that are used by the IPW, IPWR, IPWS, AIPW, and IPWREG methods to estimate the potential outcome means and ATE.

However, when you estimate the average treatment effect for the treated (ATT), the IPWR method uses the weights

\[
\frac{\Pr(T = 1 | x_i)}{\Pr(T = t_i | x_i)} = t_i + \frac{(1 - t_i)e(x_i)}{1 - e(x_i)}
\]

The differential uses of weights for ATE and ATT also apply to the following analyses or plots:

- weights that are flagged by the WGTFLAG= value
- the OUTBYWEIGHT, WDIST, and WCloud plots

Regardless of whether the ATT or ATE is estimated, the following analyses or plots always use the inverse probability weights (the first formula in this section):

- the weighted standardized mean differences and variance ratios for the covariates (effects) in the propensity score model, which are requested by the COVDIFFPS option
- the weighted density plots for the continuous covariates (effects) in the propensity score model, which are requested by the PLOTS=PSCOVDEN option

Standard Errors and Confidence Intervals

Asymptotic Formulas

PROC CAUSALTRT computes standard errors for the potential outcome means and treatment effect by using the robust sandwich covariance formula that is based on asymptotic theory. In general, let \( \theta_0 \) denote the vector of parameters of interest. PROC CAUSALTRT considers all its estimation as an M-estimation problem that solves the following estimating equations:

\[
S(\theta) = \sum_{i=1}^{n} S_i = 0
\]
When the function $S(\cdot)$ is evaluated at $\theta_0$, it satisfies $E[S] = 0$. By the theory of M-estimation (Stefanski and Boos 2002), the robust sandwich covariance matrix, $\hat{V}(\hat{\theta})$, for the estimates $\hat{\theta}$ is given by

$$n^{-1} A_n(\hat{\theta})^{-1} B_n(\hat{\theta}) A_n(\hat{\theta})^{-T}$$

where $n$ is the sample size, $A^{-T}$ denotes $(A^T)^{-1}$, and

$$A_n(\hat{\theta}) = n^{-1} \sum_{i=1}^{n} -\frac{\partial S_i}{\partial \theta}^T$$

$$B_n(\hat{\theta}) = n^{-1} \sum_{i=1}^{n} S_i S_i^T$$

PROC CAUSALTRT computes standard errors by taking the square roots of the diagonal elements of $\hat{V}(\hat{\theta})$. The form and dimension of $S(\theta)$, and hence those of the robust sandwich covariance matrix, depend on which estimation method is used. Whenever an outcome or treatment (propensity score) model is used in a particular estimation method, the estimating equations must also include the corresponding score vectors for that model.

Let $\beta$ be a generic vector of parameters of a model, and let $\mu = (\mu_0, \mu_1)$ be the vector of potential outcome means. Further, let $S_{ps}$, $S_{out}$, and $S_{ipw}$ denote the score functions from the maximum likelihood estimation of the propensity score, outcome, and inverse probability weighted outcome models, respectively. Then the composition of $\theta_0$ and $S_i$ for each estimation method is as follows:

- **AIPW**

  $\theta_0 = (\beta_{ps}, \beta_c, \beta_t, \mu)$

  $S_i = (S_{ps,i}, S_{out,i}, S_{ipw,i})$

- **IPW and IPWR**

  $\theta_0 = (\beta_{ps}, \mu)$

  $S_i = (S_{ps,i}, S_{ipw,i})$

- **IPWS**

  $\theta_0 = (\beta_{ps}, \eta, \mu)$

  $S_i = (S_{ps,i}, S_{ipw,i})$

- **IPWREG**

  $\theta_0 = (\beta_{ps}, \beta_c^w, \beta_t^w, \mu)$

  $S_i = (S_{ps,i}, S_{out,i}, S_{ipw,i})$
Chapter 36: The CAUSALTRT Procedure

- REGADJ

\[ \theta_0 = (\beta_c, \beta_t, \mu) \]

\[ S_i = (S_{out,i}, S_{reg,i}) \]

The \((1 - \alpha)100\%\) Wald confidence interval for the potential outcome mean \(\mu_i\) is given by

\[ \hat{\mu}_i \pm z_{1-\alpha/2} \hat{\sigma}_i \]

where \(z_p\) is the \(100p\)th percentile of the standard normal distribution and \(\hat{\sigma}_i\) is the standard error estimate for the potential outcome estimate \(\hat{\mu}_i\). Because the ATE is the difference in potential outcome means, its standard error is computed by applying standard variance rules to the estimates from the covariance matrix for the potential outcomes. The corresponding Wald confidence interval for ATE is then constructed by the standard formula as shown in the previous expression. For the estimates of conditional potential outcome means, the standard errors and confidence limits are computed using the same approach but with \(S_{ipw,i}\) and \(S_{reg,i}\) replaced by \(S_{ipw,i|T=1}\) and \(S_{reg,i|T=1}\).

To provide better theoretical and computational properties, PROC CAUSALTRT uses modified expressions for components of \(A_n\) and \(B_n\) when the covariance matrix \(\hat{V}(\theta)\) is computed. For descriptions of the modifications, their motivation, and theoretical justification, see Pierce (1982); Robins, Rotnitzky, and Zhao (1995); Lunceford and Davidian (2004); Wooldridge (2010).

Bootstrap Methods

If you specify the BOOTSTRAP statement, PROC CAUSALTRT uses bootstrap resampling to compute standard errors and confidence intervals for the potential outcome means and treatment effect. The procedure samples as many bootstrap sample data sets (replicates) as you specify in the NBOOT= option and then estimates the potential outcome means and treatment effect for each replication. The bootstrap samples are taken from within the treatment conditions. Each bootstrap replicate contains the same numbers of usable observations in the control and treatment conditions as the number of usable observations that are included in the input data set.

Bootstrap confidence intervals are computed only for the potential outcome means and treatment effect. You can specify one or more of the following types of bootstrap confidence intervals in the BOOTCI option in the BOOTSTRAP statement:

- The BOOTCI(NORMAL) option uses the normal approximation method to compute the bootstrap confidence interval. That is, 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. PROC CAUSALTRT requires at least 50 bootstrap samples for normal bootstrap confidence intervals and does not compute them if 40 or fewer of the samples produce usable estimates.

- The BOOTCI(PERC) option requests the percentile method, which uses the \(100(\alpha/2)\)th and \(100(1 - \alpha/2)\)th percentiles of the bootstrap parameter estimates as the confidence limits. These percentiles 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 of 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,

$$z_0 = \Phi^{-1}\left(\frac{N(\mu_j^* \leq \hat{\mu}_j)}{B}\right)$$

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.

The bias-corrected bootstrap confidence intervals are the default intervals used if you do not override them by specifying the NORMAL or PERC suboptions in the BOOTCI option.

PROC CAUSALTRT requires at least 1,000 bootstrap samples for the 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 the percentile or bias-corrected bootstrap confidence intervals are requested, the value of $n$ is ignored. Bootstrap confidence intervals and the bootstrap standard deviations are computed in the same manner for the estimates of the ATE and ATT.

### Missing Values

When the CAUSALTRT procedure fits a model, it excludes observations that have missing values for the outcome variable, treatment variable, frequency variable, or predictor variables, which are specified in the MODEL and PSMODEL statements. It also excludes observations that have invalid response or frequency values. Invalid response values are determined by the probability distribution for the response as specified in the DIST= option in the MODEL statement. The exclusion of observations because of invalid responses applies to all estimation methods, even for methods that do not require an outcome model specification (such as IPW, IPWR, and IPWS). In particular, responses must be strictly positive for the inverse Gaussian and gamma distributions, and must be nonnegative for the Poisson distribution. In all cases, exclusion of observations almost certainly affects the estimation results.
**ODS Table Names**

PROC CAUSALTRT assigns a name to each table that 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 36.8. 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</th>
</tr>
</thead>
<tbody>
<tr>
<td>CausalEffects</td>
<td>Estimates for potential outcome means and treatment effect</td>
<td>Default</td>
<td></td>
</tr>
<tr>
<td>ClassLevels</td>
<td>Classification variable levels</td>
<td>CLASS</td>
<td>Default</td>
</tr>
<tr>
<td>ControlEstimates</td>
<td>Parameter estimates for the control condition outcome model</td>
<td>PROC</td>
<td>POUTCOMEMOD</td>
</tr>
<tr>
<td>ModelInfo</td>
<td>Model information</td>
<td>Default</td>
<td></td>
</tr>
<tr>
<td>NObs</td>
<td>Number of observations</td>
<td>Default</td>
<td></td>
</tr>
<tr>
<td>PSCovDiff</td>
<td>Standardized mean differences and variance ratios of propensity score model effects</td>
<td>PROC</td>
<td>COVDIFFPS</td>
</tr>
<tr>
<td>PSMModelEstimates</td>
<td>Parameter estimates for the propensity score model</td>
<td>PROC</td>
<td>PPSMODEL</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>TreatmentEstimates</td>
<td>Parameter estimates for the treatment condition outcome model</td>
<td>PROC</td>
<td>POUTCOMEMOD</td>
</tr>
<tr>
<td>TreatmentProfile</td>
<td>Frequency counts for the treatment variable</td>
<td>Default</td>
<td></td>
</tr>
</tbody>
</table>
ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, “Statistical Graphics Using ODS.”

Before you create graphs, ODS Graphics must be enabled (for example, by specifying the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 623 in Chapter 21, “Statistical Graphics Using ODS.”

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section “A Primer on ODS Statistical Graphics” on page 622 in Chapter 21, “Statistical Graphics Using ODS.”

PROC CAUSALTRT assigns a name to each graph it creates using ODS. You can use these names to refer to the graphs when you use ODS. Table 36.9 lists the names.

To request these graphs, ODS Graphics must be enabled and you must specify the option indicated in Table 36.9.

Table 36.9  Graphs Produced by PROC CAUSALTRT

<table>
<thead>
<tr>
<th>ODS Graph Name</th>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>BootHistPlot</td>
<td>Histogram of bootstrap sample estimates</td>
<td>BOOTSTRAP</td>
<td>PLOTS=HIST</td>
</tr>
<tr>
<td>LogitPSPlot</td>
<td>Kernel density plot for the logit of the propensity score</td>
<td>PSMODEL</td>
<td>PLOTS=LOGITPSCORE</td>
</tr>
<tr>
<td>OutcomeByPScore</td>
<td>Scatter plot of outcome by propensity score</td>
<td>PSMODEL</td>
<td>PLOTS=OUTBYPSCORE</td>
</tr>
<tr>
<td>OutcomeByWeight</td>
<td>Scatter plot of outcome by weight</td>
<td>PSMODEL</td>
<td>PLOTS=OUTBYWEIGHT</td>
</tr>
<tr>
<td>PSCloud</td>
<td>Point clouds of propensity scores</td>
<td>PSMODEL</td>
<td>PLOTS=PS CLOUD</td>
</tr>
<tr>
<td>PSCovDenPlot</td>
<td>Kernel density plots of PSMODEL effects</td>
<td>PSMODEL</td>
<td>PLOTS=PSCOV DEN</td>
</tr>
<tr>
<td>PSDistPlot</td>
<td>Box plot of propensity score</td>
<td>PSMODEL</td>
<td>PLOTS=PSDIST</td>
</tr>
<tr>
<td>WeightCloud</td>
<td>Point clouds of weights</td>
<td>PSMODEL</td>
<td>PLOTS=WEIGHT CLOUD</td>
</tr>
<tr>
<td>WeightDistPlot</td>
<td>Box plot of weights</td>
<td>PSMODEL</td>
<td>PLOTS=WEIGHTDIST</td>
</tr>
</tbody>
</table>
Examples: CAUSALTRT Procedure

The following examples illustrate some of the capabilities of the CAUSALTRT procedure. These examples are not intended to represent definitive analyses of the data sets that are presented here.

Example 36.1: Estimating the ATE Using Inverse Probability Weights

This example illustrates how the CAUSALTRT procedure uses inverse probability weights to estimate the average treatment effect (ATE). The question of interest is the effect that quitting smoking has on an individual’s weight. The data for this example are a subset of data from the NHANES I Epidemiologic Follow-Up Study (NHEFS) in Hernán and Robins (2018). For the study, medical and behavioral information were collected during an initial physical examination, and again at follow-up interviews approximately one decade later.

The input data set SmokingWeight is shown in the following DATA step:

```sas
data SmokingWeight;
  input Sex Age Race Education Exercise BaseWeight Weight Change Activity YearsSmoke PerDay Quit;
  datalines;
  0 42 1 1 2 79.04 68.95 -10.09 0 29 30 0
  0 36 0 2 0 58.63 61.23 2.60 0 24 20 0
  1 56 1 2 2 56.81 66.22 9.41 0 26 20 0
  0 68 1 2 2 59.42 64.41 4.99 1 53 3 0
  0 40 0 2 1 87.09 92.08 4.99 1 19 20 0
  ... more lines ...
  0 45 0 1 0 63.05 64.41 1.36 0 29 40 0
  1 47 0 1 0 57.72 61.23 3.51 0 31 20 0
  1 51 0 3 0 62.71 . . 0 30 40 0
  0 68 0 1 1 52.39 57.15 4.76 1 46 15 0
  0 26 0 . 0 86.75 87.54 0.79 0 9 20 0
  0 29 0 2 1 90.83 106.59 15.76 1 14 30 1
;
```

The variables in the data are as follows:

- **Activity**: level of daily activity, with values 0, 1, and 2
- **Age**: age in 1971
- **BaseWeight**: weight in kilograms in 1971
- **Change**: difference in weight at the follow-up and baseline interviews (measured in kilograms)
- **Education**: level of education, with values 0, 1, 2, 3, and 4
- **Exercise**: amount of regular recreational exercise, with values 0, 1, and 2
- **PerDay**: number of cigarettes smoked per day in 1971
Example 36.1: Estimating the ATE Using Inverse Probability Weights

- Quit: 1 if an individual quit smoking between the initial and follow-up interviews; 0 otherwise
- Race: 0 for white; 1 otherwise
- Sex: 0 for male; 1 for female
- Weight: weight in kilograms at the follow-up interview
- YearsSmoke: number of years an individual has smoked

The following statements invoke PROC CAUSALTRT to estimate the average causal effect that quitting smoking has on weight gain:

```plaintext
do$ graphics on;
proc causaltrt data=smokingweight covdiffps;
   class Sex Race Education Exercise Activity Quit /desc;
   psmodel Quit = Sex Age Education Exercise Activity YearsSmoke PerDay
      /plots=(PSDist pscovden(effects(Age YearsSmoke)));
   model Change;
run;
```

The MODEL and PSMODEL statements are both required. A model for the treatment variable, Quit, is listed in the PSMODEL statement, and the outcome variable, Change, is specified in the MODEL statement. For estimation, the inverse probability weighting method with ratio adjustment (IPWR) is used by default because only a model for the treatment variable is specified. The estimation method, outcome variable, and treatment variable are all listed in the “Model Information” table in Output 36.1.1.

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Distribution</td>
</tr>
<tr>
<td>Estimation Method</td>
</tr>
<tr>
<td>Treatment Variable</td>
</tr>
<tr>
<td>Outcome Variable</td>
</tr>
</tbody>
</table>

To examine whether the weights obtained from the propensity score model balance the covariates between the treatment and control conditions, you use the COVDIFFPS option in the PROC CAUSALTRT statement to request the computation of weighted and unweighted standardized mean differences (between treatment and control conditions) and variance ratios (treatment to control) for the covariates that are specified in the PSMODEL statement. These values are displayed in the “Covariate Differences for Propensity Score Model” table in Output 36.1.2. An improvement in the covariate balance after weighting would be indicated by the following two observations when you compare the unweighted and weighted columns in the table:

- smaller standardized mean differences in the weighted column
- variance ratios closer to one in the weighted column

Establishing covariate balance under the propensity score model is an important diagnostic step. Failing to achieve covariate balance would directly challenge the assumption of strong ignorability or conditional
exchangeability, which is essential for estimating the causal effect. For more information about the assumptions and conditions required for the causal estimation methods, see the section “Causal Effects: Definitions, Assumptions, and Identification” on page 2394.

For the current example, Output 36.1.2 clearly shows that all standardized differences in the weighted column are smaller in magnitude than their counterparts in the unweighted column. All but one of the weighted standardized differences are smaller than 0.03 in magnitude. In addition, all but one of the variance ratios in the weighted column are closer to 1 than their counterparts in the unweighted column.

Notice that rows with blank values are for reference levels of the categorical covariates. The blank values indicate that they were not computed because they can be expressed in terms of the remaining levels.

**Output 36.1.2 Standardized Mean Differences**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unweighted Difference</th>
<th>Weighted Difference</th>
<th>Variance Ratio Unweighted</th>
<th>Variance Ratio Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex 1</td>
<td>-0.1603</td>
<td>-0.0200</td>
<td>0.9962</td>
<td>1.0006</td>
</tr>
<tr>
<td>Sex 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.2820</td>
<td>0.0318</td>
<td>1.0731</td>
<td>0.9847</td>
</tr>
<tr>
<td>Education 5</td>
<td>0.1660</td>
<td>0.0111</td>
<td>1.4610</td>
<td>1.0268</td>
</tr>
<tr>
<td>Education 4</td>
<td>-0.0270</td>
<td>0.0196</td>
<td>0.9167</td>
<td>1.0624</td>
</tr>
<tr>
<td>Education 3</td>
<td>-0.0472</td>
<td>-0.0015</td>
<td>0.9811</td>
<td>0.9994</td>
</tr>
<tr>
<td>Education 2</td>
<td>-0.1116</td>
<td>-0.0034</td>
<td>0.8498</td>
<td>0.9953</td>
</tr>
<tr>
<td>Education 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise 2</td>
<td>0.0568</td>
<td>-0.0029</td>
<td>1.0252</td>
<td>0.9986</td>
</tr>
<tr>
<td>Exercise 1</td>
<td>0.0398</td>
<td>0.0166</td>
<td>1.0119</td>
<td>1.0049</td>
</tr>
<tr>
<td>Exercise 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 2</td>
<td>0.0740</td>
<td>-0.0074</td>
<td>1.2182</td>
<td>0.9796</td>
</tr>
<tr>
<td>Activity 1</td>
<td>0.0268</td>
<td>0.0196</td>
<td>1.0043</td>
<td>1.0029</td>
</tr>
<tr>
<td>Activity 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YearsSmoke</td>
<td>0.1589</td>
<td>0.0253</td>
<td>1.1846</td>
<td>1.0894</td>
</tr>
<tr>
<td>PerDay</td>
<td>-0.2167</td>
<td>0.0027</td>
<td>1.1679</td>
<td>1.3323</td>
</tr>
</tbody>
</table>

Further diagnostics for assessing the covariate balance are requested by using the PLOTS= option in the PSMODEL statement.

The EFFECTS(Age YearsSmoke) suboption in the PSCOVDEN option requests that weighted and unweighted kernel density estimates be computed for the covariates Age and YearsSmoke. Output 36.1.3 displays the related paneled plots. The densities for both covariates show improved balance after weighting, as shown by the closely matched density curves in the weighted column. This observation echoes the preceding conclusion that the weights obtained from fitting the propensity score model lead to good covariate balance in Age and YearsSmoke. These density plots are available only for continuous covariates.

Output 36.1.4 shows the box plots that are requested by specifying PSDIST in the PLOTS= option. These box plots of propensity scores by treatment level indicate a similar distribution of propensity scores between treatment conditions.
The estimates for the potential outcome means and ATE are displayed in Output 36.1.5. The estimate for the ATE of 3.1876 indicates that, on average, quitting smoking leads to an increase of about three kilograms in an individual’s weight.
Output 36.1.4 Propensity Score Distributions

Output 36.1.5 IPWR Potential Outcome Means and ATE Estimates

As an alternative to estimating the ATE by using the IPWR estimation method, you can estimate it by using inverse probability weights with either the basic inverse probability weighting method (METHOD=IPW) or the inverse probability weighting method with ratio and scale adjustments (METHOD=IPWS). The following statements use the METHOD=IPWS option to estimate the ATE:

```
proc causaltrt data=smokingweight method=IPWS;
    class Sex Race Education Exercise Activity Quit /desc;
    psmodel Quit = Sex Age Education Exercise Activity YearsSmoke PerDay;
    model Change;
run;
```

Because the effects that are specified in the PSMODEL statement have not changed from the preceding analysis, the same model is fit for the treatment assignment in the current example. Therefore, the predicted weights and measures of covariate balance do not change. What has changed by using the IPWS instead of
the IPWR estimation method is how the weights are incorporated into the estimation of the potential outcome means and ATE. For this example, there is little difference between the IPWS estimate of 3.1896 for the ATE, shown in Output 36.1.6, and the IPWR estimate of 3.1876, shown in Output 36.1.5.

Output 36.1.6 IPWS Potential Outcome Means and ATE Estimates

| Parameter | Treatment Level | Estimate | Robust Std Err | Wald 95% Confidence Limits | Z | Pr > |Z||
|-----------|-----------------|----------|----------------|----------------------------|---|------|---|
| POM       | 1               | 4.9850   | 0.4530         | 4.0972                     | 5.8728 | 11.01 | <.0001 |
| POM       | 0               | 1.7954   | 0.2163         | 1.3715                     | 2.2193 | 8.30   | <.0001 |
| ATE       |                 | 3.1896   | 0.4973         | 2.2149                     | 4.1643 | 6.41   | <.0001 |

Example 36.2: Estimating the ATE Using Augmented Inverse Probability Weights

This example illustrates how the CAUSALTRT procedure uses augmented inverse probability weights (AIPW) to perform doubly robust estimation of the ATE. This example uses the SmokingWeight data set that is described in Example 36.1 and estimates the same causal effect: how quitting smoking might affect an individual’s change in weight.

The AIPW method combines modeling the treatment assignment and modeling the outcome variable to estimate the potential outcome means and the ATE. The following statements invoke PROC CAUSALTRT to estimate the causal effect of quitting smoking on weight gain by using the AIPW method:

```sas
ods graphics on;
proc causaltrt data=smokingweight method=AIPW nthreads=4;
    class Sex Race Education Exercise Activity Quit /desc;
    psmodel Quit = Sex Age Education Exercise Activity YearsSmoke PerDay ;
    model Change = Sex Age Exercise Activity BaseWeight ;
    bootstrap bootci(all) plot=hist seed=1234;
run;
```

The model for the treatment assignment is specified in the PSMODEL statement, and the outcome model is specified in the MODEL statement. The AIPW method is a doubly robust estimation method and provides unbiased estimates for the ATE even when one of the models is misspecified. The model for the treatment assignment is the same model that is used in Example 36.1. Therefore, the estimates for the weights and measures of covariate balance (Output 36.1.2 and Output 36.1.3) are the same for this example.

Although the AIPW estimates for the ATE and potential outcome means are doubly robust, the empirical estimates for their standard errors (the default estimates of standard errors) are not. You can request bootstrap-based estimation of the standard errors and confidence limits by using the BOOTSTRAP statement. The BOOTCI(ALL) option requests the estimation of all three bootstrap-based confidence intervals that PROC CAUSALTRT implements: the bias-corrected percentile method (default), percentile method, and normal approximation method. The bootstrap-based confidence intervals are added to the “Analysis of Causal Effect” table. The standard Wald confidence intervals and the three bootstrap-based confidence intervals all provide similar limits as shown in Output 36.2.1. The robust standard error estimate and the bootstrap standard error estimate are also very close. The AIPW estimate for the ATE of 3.3049 is larger than the IPWS estimate for
the ATE of 3.1896 (as shown in Output 36.1.6) and also larger than the IPWR estimate of 3.1876 (as shown in Output 36.1.5).

**Output 36.2.1 AIPW Potential Outcome Means and ATE Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Treatment Level</th>
<th>Estimate</th>
<th>Robust Std Err</th>
<th>Bootstrap Std Err</th>
<th>Wald 95% Confidence Limits</th>
<th>Bootstrap Wald 95% Confidence Limits</th>
<th>Bootstrap Percentile 95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>POM</td>
<td>1</td>
<td>5.0830</td>
<td>0.4495</td>
<td>0.4588</td>
<td>4.2019</td>
<td>5.9641</td>
<td>4.1839планета 5.9822 планета 1.907 планета 6.0400</td>
</tr>
<tr>
<td>POM</td>
<td>0</td>
<td>1.7781</td>
<td>0.2156</td>
<td>0.2227</td>
<td>1.3556</td>
<td>2.2007</td>
<td>1.3416 планета 2.2147 планета 1.3175 планета 2.2155</td>
</tr>
<tr>
<td>ATE</td>
<td></td>
<td>3.3049</td>
<td>0.4911</td>
<td>0.5016</td>
<td>2.3423</td>
<td>4.2675</td>
<td>2.3218 планета 4.2879 планета 2.3511 планета 4.3029</td>
</tr>
</tbody>
</table>

To inspect the distribution of the bootstrap estimates, you request histograms of the estimates by specifying the `PLOTS=HIST` option in the `BOOTSTRAP` statement. In **Output 36.2.2**, the histograms are collected in a panel and show no extreme estimates. Normal densities are overlaid in the histograms. **Output 36.2.2** shows that the distributions of the potential outcome means (POM) and ATE estimates are approximately normal. To reproduce the random number stream that was used to generate the bootstrap estimates, you must use the same `SEED=` value and the same number of threads for the analytic computations. When bootstrap resampling is performed, a column that indicates the number of threads used is added to the “Analysis of Causal Effect” table; this column is not displayed but is available if you use the ODS OUTPUT statement to save the table as an output data set. You can also display this column by modifying the corresponding template.
Example 36.3: Estimating Treatment Effects Using Regression Adjustment

This example demonstrates how the CAUSALTRT procedure can use regression adjustment to estimate both the ATE (average treatment effect) and the ATT (average treatment effect for the treated). The question of interest in this example is the effect that attending a Catholic high school has on a student’s performance in mathematics. The data for this example are a subset of data from the National Educational Longitudinal Study of 1988 in Murnane and Willett (2011). Demographic information was collected from the students and their families along with standardized test scores in 1988 and subsequent follow-up years. The outcome for this example is the student’s score on the mathematics portion of the standardized tests in 1992.
The input data set School is shown in the following DATA step:

```plaintext
data School;
  input Income $ FatherEd $ MotherEd $ Math BaseMath Catholic $;
datalines;
  Middle HighSchool HighSchool 49.77 50.27 Yes
  Middle Unknown Unknown 59.84 51.52 Yes
  Middle NoHighSchool Postsecondary 50.38 47.56 Yes
  Middle Unknown Unknown 45.03 46.60 Yes
  High HighSchool HighSchool 54.26 60.22 Yes
  Low NoHighSchool HighSchool 55.37 58.63 Yes
  ... more lines ...
  High SomeSecondary HighSchool 43.28 43.27 Yes
  Middle HighSchool HighSchool 41.69 48.29 Yes
  High College College 56.60 60.15 Yes
  High College SomeSecondary 56.29 59.53 Yes
  High College Postsecondary 58.16 57.06 Yes
  High College SomeSecondary 63.57 63.51 Yes
;```

The data set School consists of records for 5,671 high school students whose 1988 total family income was at most $75,000. The variables in the data are as follows:

- **BaseMath**: student’s score on the mathematics portion of a standardized test in 1988
- **Catholic**: Yes if a student attended a Catholic high school; No otherwise
- **FatherEd**: highest level of education completed by the student’s father, with six levels
- **Income**: classification based on total family income, with values of Low, Middle, and High
- **Math**: student’s score on the mathematics portion of a standardized test in 1992
- **MotherEd**: highest level of education completed by the student’s mother, with six levels

The following statements invoke the PROC CAUSALTRT procedure to estimate the ATE of attending a Catholic high school on math scores:

```plaintext
proc causaltrt data=school method=regadj poutcomemod;
  class Income FatherEd MotherEd;
  psmodel Catholic(ref='No');
  model Math = BaseMath Income FatherEd MotherEd;
run;
```

The `METHOD=REGADJ` option requests that regression adjustment be used to estimate the ATE. For this estimation method, models are fit for the outcome variable separately for each treatment condition. The treatment variable, `Catholic`, is listed in `PSMODEL` statement. The `ref='No'` option identifies `Catholic='No'` as the reference or control level. The outcome variable `Math` is specified in the `MODEL` statement along with the effects used to fit the outcome model.
The POUTCOMEMOD option in the PROC CAUSALTRT statement requests the display of the parameter estimates for the outcome model fit within each treatment condition. Two tables are produced: one for the control condition (Output 36.3.1) and one for the treatment condition (Output 36.3.2).

### Output 36.3.1 Control Group Estimates

The CAUSALTRT Procedure

<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>10.9061</td>
<td>0.4808</td>
<td>9.9638 11.8484</td>
<td>514.5891</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BaseMath</td>
<td>0.7774</td>
<td>0.0078</td>
<td>0.7622 0.7927</td>
<td>9950.9801</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Income High</td>
<td>0.4244</td>
<td>0.1818</td>
<td>0.0681 0.7807</td>
<td>5.4516</td>
<td>0.0196</td>
</tr>
<tr>
<td>Income Low</td>
<td>-0.7194</td>
<td>0.1944</td>
<td>-1.1004 -0.3385</td>
<td>13.7017</td>
<td>0.0002</td>
</tr>
<tr>
<td>Income Middle</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>FatherEd College</td>
<td>0.3857</td>
<td>0.3313</td>
<td>-0.2637 1.0351</td>
<td>1.3548</td>
<td>0.2444</td>
</tr>
<tr>
<td>FatherEd HighScho</td>
<td>-0.3112</td>
<td>0.2903</td>
<td>-0.8803 0.2579</td>
<td>1.1488</td>
<td>0.2838</td>
</tr>
<tr>
<td>FatherEd NoHighSc</td>
<td>-1.1478</td>
<td>0.3302</td>
<td>-1.7950 -0.5005</td>
<td>12.0806</td>
<td>0.0005</td>
</tr>
<tr>
<td>FatherEd Postsec</td>
<td>0.5464</td>
<td>0.3827</td>
<td>-0.2037 1.2965</td>
<td>2.0381</td>
<td>0.1534</td>
</tr>
<tr>
<td>FatherEd SomeSeco</td>
<td>0.1948</td>
<td>0.3066</td>
<td>-0.4061 0.7957</td>
<td>0.4036</td>
<td>0.5252</td>
</tr>
<tr>
<td>FatherEd Unknown</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>MotherEd College</td>
<td>0.6255</td>
<td>0.3747</td>
<td>-0.1089 1.3600</td>
<td>2.7863</td>
<td>0.0951</td>
</tr>
<tr>
<td>MotherEd HighScho</td>
<td>-0.00768</td>
<td>0.3222</td>
<td>-0.6392 0.6238</td>
<td>0.0006</td>
<td>0.9810</td>
</tr>
<tr>
<td>MotherEd NoHighSc</td>
<td>-0.5579</td>
<td>0.3610</td>
<td>-1.2654 0.1496</td>
<td>2.3890</td>
<td>0.1222</td>
</tr>
<tr>
<td>MotherEd Postsec</td>
<td>0.1472</td>
<td>0.4370</td>
<td>-0.7093 1.0036</td>
<td>0.1134</td>
<td>0.7363</td>
</tr>
<tr>
<td>MotherEd SomeSeco</td>
<td>0.4911</td>
<td>0.3346</td>
<td>-0.1647 1.1469</td>
<td>2.1540</td>
<td>0.1422</td>
</tr>
<tr>
<td>MotherEd Unknown</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
Output 36.3.2  Treatment Group Estimates

The CAUSALTRT Procedure

Outcome Model Estimates for Treatment Group

<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>15.6156</td>
<td>1.7186</td>
<td>12.2471 - 18.9840</td>
<td>82.5561</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BaseMath</td>
<td>0.7274</td>
<td>0.0244</td>
<td>0.6795 - 0.7752</td>
<td>887.3010</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Income High</td>
<td>-0.3734</td>
<td>0.5055</td>
<td>-1.3641 - 0.6173</td>
<td>0.5458</td>
<td>0.4600</td>
</tr>
<tr>
<td>Income Low</td>
<td>-1.9732</td>
<td>0.6744</td>
<td>-3.2950 - 0.6514</td>
<td>8.5605</td>
<td>0.0034</td>
</tr>
<tr>
<td>Income Middle</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FatherEd</td>
<td>2.3426</td>
<td>1.0035</td>
<td>0.3758 - 4.3095</td>
<td>5.4497</td>
<td>0.0196</td>
</tr>
<tr>
<td>FatherEd HighScho</td>
<td>0.6428</td>
<td>1.0026</td>
<td>-1.3223 - 2.6080</td>
<td>0.4111</td>
<td>0.5214</td>
</tr>
<tr>
<td>FatherEd NoHighSc</td>
<td>0.0547</td>
<td>1.0842</td>
<td>-2.0702 - 2.1796</td>
<td>0.0025</td>
<td>0.9598</td>
</tr>
<tr>
<td>FatherEd Postseco</td>
<td>1.2191</td>
<td>1.0603</td>
<td>-0.8591 - 3.2973</td>
<td>1.3219</td>
<td>0.2503</td>
</tr>
<tr>
<td>FatherEd SomeSeco</td>
<td>1.9739</td>
<td>0.9848</td>
<td>0.0436 - 3.9041</td>
<td>4.0171</td>
<td>0.0450</td>
</tr>
<tr>
<td>FatherEd Unknown</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MotherEd</td>
<td>-0.1707</td>
<td>1.0663</td>
<td>-2.2607 - 1.9193</td>
<td>0.256</td>
<td>0.8728</td>
</tr>
<tr>
<td>MotherEd HighScho</td>
<td>-1.3311</td>
<td>1.0506</td>
<td>-3.3903 - 0.7280</td>
<td>1.6053</td>
<td>0.2051</td>
</tr>
<tr>
<td>MotherEd NoHighSc</td>
<td>-0.6978</td>
<td>1.3538</td>
<td>-3.3513 - 1.9557</td>
<td>0.2656</td>
<td>0.6063</td>
</tr>
<tr>
<td>MotherEd Postseco</td>
<td>-0.8308</td>
<td>1.1261</td>
<td>-3.0379 - 1.3764</td>
<td>0.5442</td>
<td>0.4607</td>
</tr>
<tr>
<td>MotherEd SomeSeco</td>
<td>-0.8607</td>
<td>1.0016</td>
<td>-2.8238 - 1.1025</td>
<td>0.7383</td>
<td>0.3902</td>
</tr>
<tr>
<td>MotherEd Unknown</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The potential outcome means and ATE are estimated by averaging predicted values for each student from each of the outcome models. The ATE estimate of 1.6920 in Output 36.3.3 indicates that attending a Catholic high school has a positive effect on math scores. This estimate for the treatment effect is much smaller than the estimate of 3.8949 that you would obtain by comparing the unadjusted average math scores for each treatment condition.

Output 36.3.3  Regression Adjustment Potential Outcome Means and ATE Estimates

The CAUSALTRT Procedure

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Treatment Level</th>
<th>Estimate</th>
<th>Robust Std Err</th>
<th>Wald 95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>[Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>POM</td>
<td>Yes</td>
<td>52.5790</td>
<td>0.2645</td>
<td>52.0605 - 53.0975</td>
<td>198.76</td>
<td>&lt; .0001</td>
<td></td>
</tr>
<tr>
<td>POM</td>
<td>No</td>
<td>50.8871</td>
<td>0.1280</td>
<td>50.6363 - 51.1378</td>
<td>397.70</td>
<td>&lt; .0001</td>
<td></td>
</tr>
<tr>
<td>ATE</td>
<td></td>
<td>1.6920</td>
<td>0.2536</td>
<td>1.1950 - 2.1889</td>
<td>6.67</td>
<td>&lt; .0001</td>
<td></td>
</tr>
</tbody>
</table>

Conceptually, the ATE in this example is the average effect on the math scores for everyone in the population who would have attended a Catholic high school versus a non-Catholic high school. Causal effect analysis such as the current example enables you to estimate the ATE with causal interpretation, which is of general scientific interest.

However, for policy-making purposes, sometimes it is more interesting to examine the effectiveness of a program or treatment for those who did participate. The average treatment effect for the treated (ATT) is such a concept; it measures the treatment effect conditional on those who receive the treatment condition. PROC CAUSALTRT provides you two methods to estimate ATT: the inverse probability weighting method with ratio adjustment (METHOD=IPWR) or regression adjustment (METHOD=REGADJ).
The remainder of this section demonstrates the estimation of ATT for the Catholic high school data example. To estimate the average treatment effect for the treated, you specify the ATT option in the PROC CAUSALTRT statement. The following SAS statements estimate the ATT using the regression adjustment estimation method and the same outcome model as specified in the previous analysis for ATE:

```sas
proc causaltrt data=school method=regadj att;
  class Income FatherEd MotherEd;
  psmodel Catholic(ref='No');
  model Math = BaseMath Income FatherEd MotherEd;
run;
```

The “Analysis of Causal Effect” table now displays estimates for the ATT and the potential outcome means conditional on receiving the treatment condition. The estimate for the ATT is 1.5727, which is slightly smaller than the ATE estimate.

Output 36.3.4 Regression Adjustment ATT Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Treatment Level</th>
<th>Estimate</th>
<th>Robust Std Err</th>
<th>Wald 95% Confidence Limits</th>
<th>Z Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>POM</td>
<td>Yes</td>
<td>54.5395</td>
<td>0.3475</td>
<td>53.8583 55.2207</td>
<td>156.93</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>POM</td>
<td>No</td>
<td>52.9668</td>
<td>0.3034</td>
<td>52.3722 53.5614</td>
<td>174.59</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ATT</td>
<td></td>
<td>1.5727</td>
<td>0.2268</td>
<td>1.1281 2.0173</td>
<td>6.93</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

From a policy-making point of view, the average effect that attending a Catholic high school would have on students who are in non-Catholic high school would also be of interest. Such a mirror concept of ATT is also known as ATU, average treatment effect for the untreated. Although the CAUSALTRT procedure does not provide a direct option for estimating ATU, you can still achieve that purpose by using the ATT option with the reverse treatment designation. In the following PROC CAUSALTRT statements, the ref='Yes' option in the PSMODEL statement redefines the treatment condition to be attending a non-Catholic high school.

```sas
proc causaltrt data=school method=regadj att;
  class Income FatherEd MotherEd;
  psmodel Catholic(ref='Yes');
  model Math = BaseMath Income FatherEd MotherEd;
run;
```

The ATT estimate of –1.7059 in Output 36.3.5 is the average causal effect that attending non-Catholic high schools has on the math scores of students in non-Catholic high schools. Reversing the sign of this estimate provides an estimate for the ATU: the average treatment effect that attending a Catholic school would have for those students in non-Catholic high schools. Hence, the ATU in the current example is 1.7059.

In some research studies, researchers would like to compare ATT and ATU to investigate the so-called effect heterogeneity issue. In the current example, the ATU is only slightly larger (by about 0.13 points) than the ATT. Considering the fact that all potential outcome means are about 50, it thus appears to be safe to say that effect heterogeneity, as indicated by the 0.13 point difference, is not an issue in the current example.
Output 36.3.5  AIPW Potential Outcome Means and ATE Estimates

The CAUSALTRT Procedure

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Treatment Level</th>
<th>Estimate</th>
<th>Robust Std Err</th>
<th>Wald 95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>POM</td>
<td>No</td>
<td>50.6447</td>
<td>0.1338</td>
<td>50.3825 50.9068</td>
<td>378.60</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>POM</td>
<td>Yes</td>
<td>52.3505</td>
<td>0.2717</td>
<td>51.8179 52.8831</td>
<td>192.66</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td></td>
<td>-1.7059</td>
<td>0.2608</td>
<td>-2.2171 -1.1947</td>
<td>-6.54</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

References


MIT Press.
Subject Index

AIPW estimates
  CAUSALTRT procedure, 2401
alpha level
  CAUSALTRT procedure, 2373
ATE
  CAUSALTRT procedure, 2396
ATT
  CAUSALTRT procedure, 2396
average treatment effect (ATE)
  CAUSALTRT procedure, 2396
average treatment effect for the treated (ATT)
  CAUSALTRT procedure, 2396
bias-corrected confidence interval
  CAUSALTRT procedure, 2381, 2407
bootstrap normal confidence interval
  CAUSALTRT procedure, 2406

CAUSALTRT procedure
  AIPW estimates, 2401
  ATE, 2396
  ATT, 2396
  average treatment effect (ATE), 2396
  average treatment effect for the treated (ATT), 2396
  bias-corrected confidence interval, 2381, 2407
  bootstrap normal confidence interval, 2406
  conditional exchangeability, 2396
  confidence interval, 2381
  counterfactual conditionals, 2395
  default estimation methods, 2367
  inverse probability weights, 2398
  IPW estimates, 2398
  IPWR ATT estimates, 2403
  IPWR estimates, 2399
  IPWREG estimates, 2402
  IPWS estimates, 2399
  no hidden variations of treatments, 2395
  no interference, 2395
  no unmeasured confounders, 2398
  ODS table names, 2408
  ordering of effects, 2377
  outcome level ordering, 2386
  output ODS Graphics table names, 2409
  percentile confidence interval, 2407
  positivity assumption, 2398
  potential outcomes, 2395
  propensity score, 2397
  REGADJ ATT estimates, 2403
  robust sandwich variance estimate, 2405
  stable unit treatment value assumption (SUTVA), 2395
  strong ignorability, 2396
  SUTVA, 2395
  treatment level ordering, 2391
  treatment value irrelevance, 2395
classification variables
  sort order of levels (CAUSALTRT), 2377
conditional exchangeability
  CAUSALTRT procedure, 2396
confidence interval
  CAUSALTRT procedure, 2381
  counterfactual conditionals
    CAUSALTRT procedure, 2395
influence probability weights
  CAUSALTRT procedure, 2398
IPW estimates
  CAUSALTRT procedure, 2398
IPWR ATT estimates
  CAUSALTRT procedure, 2403
IPWR estimates
  CAUSALTRT procedure, 2399
IPWREG estimates
  CAUSALTRT procedure, 2402
IPWS estimates
  CAUSALTRT procedure, 2399
no hidden variations of treatments
  CAUSALTRT procedure, 2395
no interference
  CAUSALTRT procedure, 2395
no unmeasured confounders
  CAUSALTRT procedure, 2398
outcome level ordering
  CAUSALTRT procedure, 2386
output ODS Graphics table names
  CAUSALTRT procedure, 2409
percentile confidence interval
  CAUSALTRT procedure, 2407
positivity assumption
  CAUSALTRT procedure, 2398
potential outcomes
  CAUSALTRT procedure, 2395
probability distribution, built-in
propensity score
CAUSALTRT procedure, 2397

REGADJ ATT estimates
CAUSALTRT procedure, 2403
REGADJ estimates
CAUSALTRT procedure, 2400
robust sandwich variance estimate
CAUSALTRT procedure, 2405

stable unit treatment value assumption (SUTVA)
CAUSALTRT procedure, 2395
strong ignorability
CAUSALTRT procedure, 2396
SUTVA
CAUSALTRT procedure, 2395

treatment level ordering
CAUSALTRT procedure, 2391
treatment variation irrelevance
CAUSALTRT procedure, 2395
Syntax Index

ABSCONV option
  PROC CAUSALTRT statement, 2375
ABSFCONV option
  PROC CAUSALTRT statement, 2375
ABSGCONV option
  PROC CAUSALTRT statement, 2375
ALL option
  PROC CAUSALTRT statement, 2378
ALPHA= option
  CAUSALTRT procedure, MODEL statement, 2388
  CAUSALTRT procedure, PSMODEL statement, 2392
  PROC CAUSALTRT statement, 2373
ATT option
  PROC CAUSALTRT statement, 2374
BOOTCI option
  BOOTSTRAP statement (CAUSALTRT), 2381
BOOTDATA option
  BOOTSTRAP statement (CAUSALTRT), 2382
BOOTSTRAP statement
  CAUSALTRT procedure, 2381
BY statement
  CAUSALTRT procedure, 2384
CAUSALTRT procedure
  syntax, 2372
CAUSALTRT procedure, BOOTSTRAP statement, 2381
  BOOTCI option, 2381
  BOOTDATA option, 2382
  NBOOT option, 2382
  NOSKIP option, 2382
  PLOTS option, 2382
  SEED option, 2383
CAUSALTRT procedure, BY statement, 2384
CAUSALTRT procedure, CLASS statement, 2384
  CPREFIX= option, 2384
  DESCENDING option, 2385
  LPREFIX= option, 2385
  MISSING option, 2385
  ORDER= option, 2385
  REF= option, 2385
  TRUNCATE option, 2386
CAUSALTRT procedure, FREQ statement, 2386
CAUSALTRT procedure, MODEL statement, 2386
  ALPHA= option, 2388
  DESCENDING option, 2386
CAUSALTRT procedure, OUTCOME statement, 2386
CAUSALTRT procedure, OUTPUT statement, 2389
  keyword= option, 2390
  OUT= option, 2390
CAUSALTRT procedure, PROC CAUSALTRT statement, 2373
  ABSCONV= option, 2375
  ABSFCONV= option, 2375
  ABSGCONV= option, 2375
  ALL option, 2378
  ALPHA= option, 2373
  ATT option, 2374
  COVDIFFPS option, 2374
  DATA= option, 2374
  FCONV= option, 2376
  GCONV= option, 2376
  MAXFUNC= option, 2376
  MAXITER= option, 2376
  MAXTIME= option, 2377
  METHOD= option, 2374
  NAMELEN= option, 2375
  NLOPTIONS option, 2375
  NOEFFECT option, 2377
  NOPRINT option, 2377
  NTHREADS= option, 2381
  ORDER= option, 2377
  PALL option, 2378
  PLOTS= option, 2378
  POUTCOMEMOD option, 2380
  PPSMODEL option, 2380
  PREGADJ option, 2380
  PTREATMOD option, 2380
  RORDER= option, 2380
  TECHNIQUE= option, 2377
  THREADS= option, 2381
CAUSALTRT procedure, PSMODEL statement, 2390
  ALPHA= option, 2392
  DESCENDING option, 2391
  EVENT= option, 2391

DIST= option, 2388
DISTRIBUTION= option, 2388
EVENT= option, 2386
LINK= option, 2388
NOSCALE option, 2389
ORDER= option, 2387
REFERENCE= option, 2387
SCALE= option, 2389
ORDER= option, 2391
PLOTS option, 2392
REFERENCE= option, 2392
WGTFLAG= option, 2394
CAUSALTRT procedure, TREATMENT statement, 2390

CLASS statement
CAUSALTRT procedure, 2384
COVDIFFPS option
PROC CAUSALTRT statement, 2374
CPREFIX= option
CLASS statement (CAUSALTRT), 2384

DATA= option
PROC CAUSALTRT statement, 2374
DESCENDING option
CLASS statement (CAUSALTRT), 2385
MODEL statement
CAUSALTRT procedure, 2386
NAMELEN= option
PROC CAUSALTRT statement, 2375

DIST= option
MODEL statement, 2388
DISTRIBUTION= option
MODEL statement, 2388

EVENT= option
MODEL statement, 2386
PSMODEL statement, 2391

FCONV option
PROC CAUSALTRT statement, 2376
FREQ statement
CAUSALTRT procedure, 2386

GCONV option
PROC CAUSALTRT statement, 2376

keyword= option
OUTPUT statement (CAUSALTRT), 2390

LINK= option
MODEL statement, 2388
LPREFIX= option
CLASS statement (CAUSALTRT), 2385

MAXFUNC= option
PROC CAUSALTRT statement, 2376
MAXITER= option
PROC CAUSALTRT statement, 2376
MAXTIME= option
PROC CAUSALTRT statement, 2377
METHOD= option
PROC CAUSALTRT statement, 2374
MISSING option
CLASS statement (CAUSALTRT), 2385

MODEL statement
CAUSALTRT procedure, 2386

NAMELEN= option
PROC CAUSALTRT statement, 2375

NBOOT option
BOOTSTRAP statement (CAUSALTRT), 2382
NLOPTIIONS option
PROC CAUSALTRT statement, 2375
NOEFFECT option
PROC CAUSALTRT statement, 2377
NOPRINT option
PROC CAUSALTRT statement, 2377

NOSCALE option
MODEL statement, 2389
NOSKIP option
BOOTSTRAP statement (CAUSALTRT), 2382

NTHREADS= option
PROC CAUSALTRT statement, 2381

ORDER= option
CLASS statement (CAUSALTRT), 2385
MODEL statement, 2387
PROC CAUSALTRT statement, 2377
PSMODEL statement, 2391

OUT= option
OUTPUT statement (CAUSALTRT), 2390
OUTCOME statement
CAUSALTRT procedure, 2386
OUTPUT statement
CAUSALTRT procedure, 2389

PALL option
PROC CAUSALTRT statement, 2378
PLOTS option
BOOTSTRAP statement (CAUSALTRT), 2382
PSMODEL statement (CAUSALTRT), 2392

PLOTS= option
PROC CAUSALTRT statement, 2378

POUTCOMEMOD option
PROC CAUSALTRT statement, 2380

PSMODEL option
PROC CAUSALTRT statement, 2380

PREGADJ option
PROC CAUSALTRT statement, 2380

PTREATMOD option
PROC CAUSALTRT statement, 2390

REF= option
CLASS statement (CAUSALTRT), 2385
REFERENCE= option
MODEL statement, 2387
PSMODEL statement, 2392
RORDER= option
   PROC CAUSALTRT statement, 2380
SCALE= option
   MODEL statement, 2389
SEED option
   BOOTSTRAP statement (CAUSALTRT), 2383
TECHNIQUE= option
   PROC CAUSALTRT statement, 2377
THREADS= option
   PROC CAUSALTRT statement, 2381
TREATMENT statement
   CAUSALTRT procedure, 2390
TRUNCATE option
   CLASS statement (CAUSALTRT), 2386
WGTFLAG= option
   PSMODEL statement, 2394