PROCESS Documentation Addendum

PROCESS is documented in Appendices A and B of the second edition of *Introduction to Mediation, Moderation, and Conditional Process Analysis*. This addendum to the documentation describes options and features added to PROCESS since the first printing of the book in December 2017. The most recent version of this addendum was produced on August 8, 2019.

Dichotomous *Y*

(Added in version 3.1)

With the release of version 3.1, PROCESS will accept a dichotomous outcome variable Y. No input is needed from the user to specify that Y is dichotomous; PROCESS figures this out on its own. When PROCESS sees only two values in the variable specified as Y, the estimation of Y is conducted using logistic regression, modeling the probability of the event coded in Y, with the event being modeled represented in the data with the largest numerical code in the variable specified as Y. For example, if Y is coded 0 and 1, the probability of $Y = 1$ is estimated, and all effects should be interpreted accordingly. All regression coefficients for the model of Y are logistic regression coefficients and are on a log-odds metric, including confidence intervals for these coefficients. These can be exponentiated to yield effects on an odds ratio metric.

Inferences for each regression coefficient in the model of Y are based on ratio of the estimate of the coefficient to its standard error. P-values and confidence intervals are based on the assumption of a normally distributed sampling distribution of the regression coefficient. Hence, the hypothesis test that a regression coefficient is equal to zero is equivalent to a Wald test. Inference for the complete model of Y is based a likelihood ratio test with degrees of freedom equal to the number of predictors in the model of Y.

When the model of Y includes a moderation component, the section of output for the model of Y labelled "test for highest-order unconditional interaction(s)" is based on a likelihood ratio test, comparing the fit of the model of Y that includes the interaction compared to a model that excludes it. For single-degree-of-freedom tests of interactions, this test can produce a p-value for the interaction that is different than the p-value produced for the ratio of the regression coefficient to its standard error, as likelihood ratio and Wald tests are equivalent only asymptotically and can produce different results in finite samples.

Although likelihood ratio tests are used for tests of unconditional interactions, tests of conditional interactions as well as tests of conditional effects of X when X is specified as multicategorical are Wald tests.

When the **plot** option is used with a dichotomous Y, estimates of Y for combinations of focal predictor and moderator(s) are estimated log odds of the event for those combinations. PROCESS also produces the estimated probability of the event under the label "prob."

Indirect effects of X when Y is dichotomous are calculated as always as the product of the effect of X on the mediator M and the effect of mediator M on Y controlling for X. The indirect effect (and direct effect of X) is on a log-odds metric given that the effect of the mediator on Y is a logistic regression coefficient. Because the regression coefficient for X in a model of Y without the mediators included is not equal to the sum of the direct and indirect effects of X, the **total** option is not available with a dichotomous Y. The **effsize** option is also disabled for models with a dichotomous Y.

Logistic regression is computationally more intensive than ordinary least squares regression analysis. When combined with repeated model estimation using bootstrapping for inference about indirect effects, time to produce output can be considerable even in modest sample sizes. Expect PROCESS to take longer to show results than when Y is not dichotomous. A good strategy when estimating a model with indirect effects is to first set the number of bootstrap samples to a smaller number than the default of 5000 (e.g., **boot=1000**), even zero, to make sure the model will estimate and PROCESS will produce output. Once you are satisfied, increase the number of bootstrap samples to generate the final result that you interpret.

In addition, the logistic regression estimation module programmed in PROCESS is more susceptible to matrix inversion errors and also errors produced by "perfect separation" (as it is called in the logistic regression literature). These errors are more likely to occur when X and/or moderators W or Z are multicategorical or when the event is rare or very frequent in the data. If you see INV, EXP, or other errors in the SPSS output or SAS log file, do not interpret the output.

The user has control over the maximum number of iterations before PROCESS provides a solution as well as the convergence criterion. The convergence criterion is met when changes to the logistic regression coefficients result in a change in the likelihood of less than the specified criterion. At that point, the iteration stops and the solution is provided. A failure to converge can sometimes be rectified by increasing the number of iterations or increasing the convergence criterion. However, doing so will increase computational time, perhaps dramatically. This can be done with the **iterate** and **converge** options. The defaults are 100 and .0001, respectively (i.e., **iterate=100** and **converge=.0001**).

Specifying a Moderator as a Covariate

(Added in version 3.2)

On pages 630-632 of *Introduction to Mediation, Moderation, and Conditional Process Analysis*, a trick is described for how to specify a moderator in one equation as a covariate in another equation. With the release of PROCESS v3.2, this trick is no longer necessary for a dichotomous or continuous moderator. PROCESS now allows a moderator (a variable specified as *W* or *Z*) to be specified as a covariate in the usual way by listing it following **cov=**. By default, such a variable will be added to all model equations except those where it already plays the role of a moderator variable. To include it in only some of the equations, use the **cmatrix** option described in Appendix A.

For example, to estimate the model depicted in Figure B.7 (see page 631), the SPSS PROCESS command is

```
process y=tile/m=wine/x=baby/w=milk/cov=sand tent milk/model=7.
```

In SAS, the equivalent command is

```
%process (data=four,y=tile,m=wine,x=baby,w=milk,cov=sand tent milk,
   model=7);
```

Note that this does not work for a moderator specified as multicategorical using the **mcw** or **mcz** options. To specify a multicategorical moderator in one equation as a covariate in another, the trick on pages 630-632 must still be used after first manually constructing a set of codes (e.g., indicator, Helmert, etc.) representing the multicategorical variable and adding them to the data file.

Standardized Regression Coefficients

(Added in version 3.2)

In mediation-only models (i.e., models with no product terms to capture moderation), the **effsize** option produces the completely and partially standardized direct and indirect effects in the summary section of the PROCESS output (see section 4.3 of *Introduction to Mediation, Moderation, and Conditional Process Analysis*). However, the regression coefficients elsewhere in the output are printed in unstandardized form. The release of version 3.2 adds an option for generating standardized regression coefficients for the regression models of mediators and *Y* provided in the PROCESS output. To generate standardized regression coefficients, add **stand=1** to the PROCESS command. These regression coefficients will be in completely standardized form. But because a completely standardized regression coefficient is not meaningful for a dichotomous or multicategorical variable, if the variable specified as *X* in the PROCESS command is dichotomous or multicategorical, then standardized regression coefficients for *X* (or multicategorical codes representing *X*) will be in partially standardized form. A partially standardized regression coefficient is equal to the unstandardized regression coefficient divided by the standard deviation of the outcome variable (i.e., the variable on the left side of the regression equation). That is, unlike a completely standardized regression coefficient, which removes the metrics of the predictor and the outcome from the scaling of the unstandardized regression coefficient, the partially standardized regression coefficient retains the scaling of *X*. See the discussion about partially and completely standardized regression coefficients on pages 52-54 of *Introduction to Mediation, Moderation, and Conditional Process Analysis*.

Standardized regression coefficients are not available in models that include any products of variables to represent an interaction. Standardized regression coefficients are also not available when Y is dichotomous. Note that standardized regression coefficients for dichotomous covariates (i.e., variables listed in the **`cov`** option) will be in completely standardized form. These should not be interpreted or reported.

New Mean Centering Option

(Added in version 3.4)

The **`center=1`** option available in all earlier releases of PROCESS mean centers continuous and dichotomous variables that are used to construct products of variables when parameterizing an interaction between those variables (codes representing a multicategorical focal predictor or moderator are not centered when the **`center`** option is used). With the release of version 3.4, a new **`center=2`** option mean centers only the continuous variables used to form a product. A dichotomous focal predictor or moderator used in the construction of a product is kept in its uncentered form when this option is used. Therefore, if all focal predictors and moderators are dichotomous, the **`center=2`** option has no effect.

Testing for Interaction Between X and a Mediator

(Added in version 3.4)

An indirect effect of X on Y is a product of at least two effects, such as the effect of X on mediator M and the effect of M on Y when X is held constant. The effect of M is estimated assuming that this effect does not depend on X, the assumption of *no X by M interaction*. PROCESS cannot estimate any model that allows X to moderate any path in a mediation model. However, with the release of version 3.4, PROCESS now provides a test of this no-interaction assumption. By specifying **`xmtest=1`** in the PROCESS command line, PROCESS will conduct a test of interaction between X and each variable specified as a mediator in models of outcomes causally downstream.

Depending on the complexity of the model, there may be many such tests conducted. For example, in a parallel multiple mediator model with four mediators, there will be four such tests, one for each mediator in the model of Y. In a serial multiple mediator model with two mediators, there will be three tests, one for each mediator in the model of Y, and one for M_1 in the model of M_2.

This test is conducted by assuming no interaction between X and any mediator in the model specified, and then releasing this constraint for each of the paths leading from a mediator to another variable, while all other effects of the other mediators in the model are assumed to be independent of X. For tests involving a continuous Y, the test takes the form of an F-ratio when comparing the fit of the constrained and unconstrained model. For tests involving a dichotomous Y, a likelihood ratio test is conducted.

Consider, for example, the output below generated for the parallel multiple model discussed in Chapter 5 by adding **xmtest=1** to the PROCESS command line on page 156:

```
Test(s) of X by M interaction:
              F        df1        df2          p
M1*X      .0110     1.0000   118.0000      .9166
M2*X     2.2035     1.0000   118.0000      .1404
```

The first row indicates no statistically significant interaction between experimental condition and topic importance in the model of reaction to the story, $F(1, 118) = 0.011, p = .917$. This test is conducted while fixing the effect of presumed media influence (M_2) to be unmoderated by experimental condition. The second row indicates no interaction between presumed media influence and experimental condition in this same model, $F(1, 118) = 2.204, p = 0.140$, fixing the effect of topic importance (M_1) to be independent of experimental condition. Combined, these tests support the assumption of no interaction between X and either mediator in this parallel multiple mediator model.

Note that this test is not provided for a mediator if the model specifies a three way interaction between that mediator and moderators W and Z (as in, for example, model 18). This test is suppressed in this circumstance because the test of interaction would be conditioned on holding either W or Z (or both) fixed at zero, which may not be meaningful or substantively interpretable.

If evidence of interaction between X and a mediator is found, that interaction can be probed by estimating an equivalent moderation-only model such as model 1 with X specified as the moderator W and M specified as the focal predictor X. The larger the difference in the effect of M across values of X, the less meaningful will be the indirect effect of X. Judgment must be made by the investigator as to whether such interaction jeopardizes the meaningfulness or validity of an inference from a mediation or conditional process analysis. There are methods for mediation analysis that allow for X by M interaction, though these methods are restricted in application to relatively simple models. See section 14.6 for a discussion.

Joint Test of Additive Moderation

(Added in version 3.4)

When an effect in a model is specified as moderated additively by two moderators W and Z (e.g., in model 2, the effect of X on Y; in model 16, the path from M to Y), PROCESS now provides a joint test of the null hypothesis that neither W nor Z moderates the effect. When the outcome variable (i.e., the variable on the left side of the equation) is continuous, this test takes the form of an F-ratio. When it is dichotomous, a likelihood ratio test is provided.

Consider, for example, the section of output below from model 2, which specifies that the effect of X on Y is moderated additively by both W and Z. The joint test is listed in the row labeled "Both."

```
Test(s) of highest order unconditional interaction(s):
        R2-chng          F        df1        df2          p
X*W       .0640    14.2636     1.0000   202.0000      .0002
X*Z       .0301     6.7192     1.0000   202.0000      .0102
BOTH      .0697     7.7643     2.0000   202.0000      .0006
```

In this example, we can reject the null hypothesis that neither W nor Z moderates the effect of X on M, $F(2, 202) = 7.764, p < .001$. Adding the two products to the model allowing the effect of X on M to be linearly moderated by W and Z increases the squared multiple correlation by .070.