

Interaction Macro Extension

* This is an extension of the MODPROBE macro (Hayes, 2009) for probing single-df interactions in OLS and logistic regression models. The extension (Hofmann, 2010) includes a variety of options for graphical inspection, uni- and multivariate outlier detection, and probing.

* Original macro written by Andrew F. Hayes, School of Communication, The Ohio State University. Version 1.2, 1 Sept 2009.

* Extended by Wilhelm Hofmann, University of Würzburg, Germany, hofmannw@psychologie.uni-wuerzburg.de. Version 2.5, April 2010.

Short overview of Extension Features

- Request and probe unstandardized, half-standardized, or fully-standardized line graph of interaction (ready to be transferred to word processing file)
- Plot either predicted log-odds or probabilities for logistic regression analysis
- Test and plot at different values of the moderator and specify any range of the predictor variable for the plot
- Conduct univariate and/or multivariate outlier exclusion using a variety of measures and descriptive screening options
- Check for normality
- Include additional filter variable
- do all this with just one syntax line

Preparatory Steps

To make this macro work properly, insert and run the following commands in your ordinary syntax prior to running the macro:/.

- 1) Copy the macro file 'modprobe_extension.sps' and the template 'modtemp.sgt' into your working folder.
- 2) Open the macro .sps file in SPSS. In order to activate the macro, **select and run the macro syntax page once** [pressing Ctrl+A and then Ctrl+R]. Note: using the include command instead does not work properly with this macro.
- 3) **Create a new syntax file** for your moderator analyses.
- 4) In this syntax file, **designate the working folder** by using the command line `cd 'D:\folder\folder\folder ... '.`
- 5) Now **get your .sav data** by using the get file command.
`get file=data.sav. [insert name of your datafile to get your data].`
- 6) Use the following command (to give the current dataset a name - important for the macro to run properly):
`dataset name data.`

Thus, the first three lines of your new syntax file should look like this (using example data):

```
cd 'D:\my data\study_1'.  
get file=data04.sav.  
dataset name data.
```

Short Instructions on Macro Command

Essential elements of the macro command

MODPOBE y=[name of dependent variable]/x = [your focal predictor variable] /m = [the moderator variable] /cov = [list of covariates separated by spaces].

Note that the structure of these command lines have been changed as to avoid possible confusion in the assignment of variables to focalvariable (x), moderator (m) and covariates (cov).

Optional original extensions (see Hayes & Matthes 2009).

Note: The following command elements can be inserted at any position in the above command line, ideally after the essential elements.

/modval = request a conditional effect (see also **m**low and **m**high function below)

/jn = 1 request Johnson-Neyman technique

/alpha = ... specify level of significance for Johnson-Neyman technique

/iterate = ... number of iterations

/converge = ... convergence criterion

New modifications and extensions (Hofmann, 2010)

Probing the interaction

I) Specify table and graphical output

/est = **0** only show regression output

1 add graphical interaction plot

2 show table of predicted values saved to estvals.sav.

note: style of plot can be changed by modifying the template 'modtemp.sgt'.

II) Specify solution (affecting table and graphical output)

/center = **0** unstandardized solution, all vars in original metric

1 centered solution; x and m centered around mean, y in original metric (default)

2 half-standardized solution; x and m are z-standardized, y left in original metric

3 half-standardized solution, x and m z-standardized, values of x range from -1 to 1

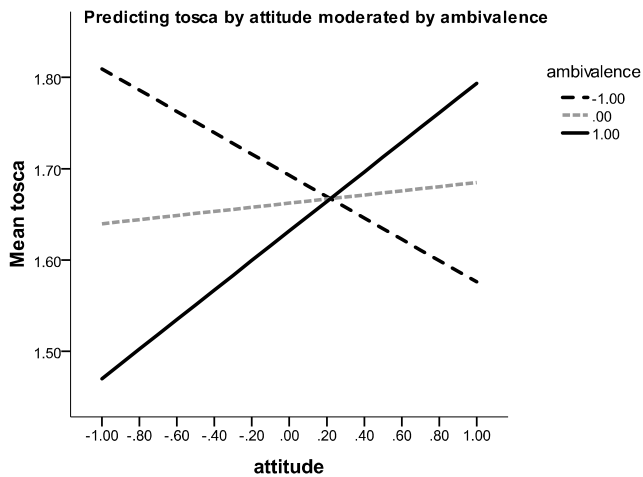
4 fully standardized solution; x, m, and y are z-standardized; across full range of x

5 fully standardized solution, x, m, and y are z-standardized, x ranges from -1 to 1

Note: covariates are always centered around mean; standard values for m are -1 SD, 0 (the mean if centered), and +1 SD

Example for center = 3 (x and m standardized, x ranging from -1SD to 1SD, y unstandardized) using modtemp.sgt as the template:

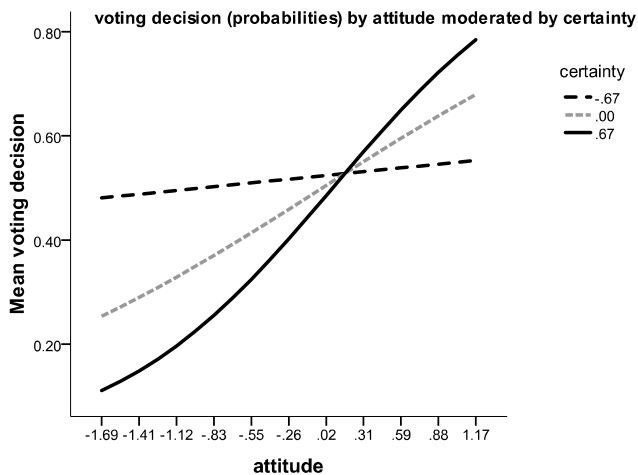
modprobe y=tosca/x=attitude/m=ambivalence/center=3.



III) For logistic regression, choose between log-odds and probabilities plot
/plotprob = 0 in logistic regression, plot predicted log-odds on y-axis (default)
1 plot predicted probabilities on y-axis
2 plot predicted odds on y-axis

Example for center = 1 (x and m mean-centered) and plotprob = 1 using modtemp.sgt as the template:

modprobe y=voting/x=attitude/m=certainty/center=1/plotprob=1.



Note: plotprob = 1 produces an error message if combined with a continuous dependent variable

IV) Customize x and m values (in SD units) at which interaction is probed
/xlow = ... /x high = ... [number of SDs (e.g., -1 and 1) from sample mean for which you want plot (x-axis)/predictions; default of the macro is to use full observed range of x variable].
 Note that xlow and xhigh is overruled and set to +/-1 if center = 3 or center = 5.

/m low = ... /m high = ... [number of SDs (e.g., -1 and 1) from sample mean for which you want conditional testing and plot/predictions]

Univariate Outlier Treatment

Note: The macro produces an outlier summary showing how many cases were excluded due to the specified (univariate and/or multivariate) outlier treatment. Outliers are not deleted from dataset.

/boxplot = 1 [displays boxplot of y,x,m, and cov for univariate outlier inspection using untransformed variables]
2 [displays boxplot of y,x,m, displaying z-transformed variables]

/yout = ... /xout = ... /mout = ... [specifies univariate outlier thresholds separately for y, x, and m; e.g. yout = 3 excludes cases with outliers above or below 3 SD on dependent variable]

/allout = ... [specifies ONE threshold for all univariate outlier analyses – recommended; this command overrules any yout, xout, and mout command line]

Multivariate Outlier Treatment

Note: The macro produces an outlier summary showing how many cases were excluded due to the specified (univariate and/or multivariate) outlier treatment. Outliers are not deleted from dataset.

/scatter = 1 [displays scattergram matrix for y, x, and m variables useful to spot bivariate outliers graphically]
2 [same as scatter = 1, creating a three-dimensional scattergram]

/studres = ... /sdres = ... /mahal = ... /lever = ... /dffits = ... /cook = ...
specifies (absolute) multivariate outlier thresholds for the 6 respective statistics (if a statistic is not specified, it is not in effect). Macro will report a table with an overview of these (and other) residual statistics.

Note that Multivariate Outliers can be computed in a variety of ways and that there is no definite approach (especially, on which cutoff values to use). The macro includes six commonly used options (studres, sdres, mahal, lever, dffits, cook). Histograms (**histo**) of the distribution on these statistics can help to visually “spot” multivariate outliers. You can individually specify the cutoff values you want to use (e.g., /mahal = 20). In case you are unfamiliar with these measures and suggested cutoffs, the macro can suggest cutoff values according to established conventions (see below) by specifying “9999” as the cutoff value (e.g., /cook = 9999). The macro then calculates, displays, and uses these cutoff values for multivariate outlier exclusion. Finally, you can choose whether you want to exclude cases that exceed the cutoff on at least one of the statistics employed (**mtreat = OR**) or that exceed all cutoff values simultaneously (**mtreat = AND**).

1) **studres**: The studentized residual (or “internally” studentized residual), a measure of how much the predicted value for a given data point deviates from the actual values. Because studres (and sdres) are distributed around zero, values exceeding the specified absolute threshold either in the positive or negative direction are excluded.

Convention (**/studres = 9999**): activates the convention according to which studentized residuals exceeding the (absolute) value of 3 are deemed multivariate outliers (Cohen et al., 2003).

2) **sdres**: The deleted studentized residual (also known as “externally” studentized residual); similar to studres, except that the regression parameters are first calculated without the datapoint in question (“deleted”) and then the studentized residual is estimated for that datapoint based on these parameters.

Convention (**/sdres = 9999**): employs the convention according to which studentized deleted residuals exceeding the (absolute) value of 3 are deemed multivariate outliers (Cohen et al., 2003).

3) **mahal**: This is the squared Mahalanobis distance score. Higher scores indicate greater deviation from the mean of the distribution.

Convention (**/mahal = 9999**): the convention with Mahalanobis distance is based on the chi-square distribution with $df = k$ (number of predictors). Cases exceeding the critical value for $p = .001$ in the distribution are considered multivariate outliers. By specifying “9999”, the macro automatically calculates and uses this value as the cutoff.

4) **lever**: Centered Leverage Values. Can be transformed into Mahalanobis distance scores. Centered Leverage indicates how far a data point is from the centroid of the predictor space. Higher scores indicate greater leverage.

Convention (**/lever = 9999**): This follows the convention of Belsey et al. (1980) according to which leverage values above $3 \cdot k/N$ (k being the number of predictors, N the sample size) indicate multivariate outliers. $3k/N$ is the less liberal cutoff (Belsey et al. also recommend $2k/N$ for large samples) but may still identify too many cases as outliers (Cohen et al., 2003).

5) **dffits**: is a measure of global influence. DFFITS stands for “difference in fit, standardized” and is defined as the change (“DFFIT”), in the predicted value for a point, obtained when that point is left out of the regression (dffit), and normed by dividing it by the estimated standard deviation of the fit at that point (sdffit). DFFITS is zero when the deletion of a case does not change the regression line. Greater positive or negative values indicate greater deviations.

Convention (**/dffits = 9999**): the macro follows Cohen et al. (2003) who suggest to investigate conspicuous values exceeding $\left| 2 \sqrt{(k+1)/N} \right|$.

6) **cook**: Cook’s distance is another commonly used estimate of the global influence of a data point in regression. It is closely related to dffits and ranges upward from the minimal value of zero.

Convention (**/cook = 9999**): Cook recommends to use the 50% quantile of the F-distribution with $df_1 = k+1$ and $df_2 = n-k-1$. By specifying “9999”, the macro automatically calculates and uses this value as the cutoff. An alternative convention that is often used is to specify a value of 1 as the threshold (i.e., $cook = 1$).

/histo = 1 requests histograms for all multivariate outlier statistics specified; this is strongly recommended when multivariate outliers are excluded, since graphical inspection often allows to detect conspicuous values

/mtreat = 0 multivariate outlier statistic thresholds are implemented in an OR fashion, i.e., exceeding threshold on either criterion leads to exclusion – this is the default
1 thresholds are implemented in an AND fashion, i.e., only excess on all criteria leads to exclusion

Additional add-ons

/norm = 1 [requests a detrended normal P-P Plot in order to check for deviations from normality on all variables]

/corr = 1 requests a bivariate correlation table of y, x, m, and the covariates

addfilter = ... name of an additional pre-filter variable (e.g., 'native speakers') in dataset where a value of 1 on the filter variable indicates 'inclusion' and a value of 0 'exclusion'; no other values are possible

deletevars = 0 do not delete newly calculated variables (e.g., outlier cutoffs) computed by macro
1 delete new variables after each run (default)

Examples

The following examples can be run using `examples.sps` as syntax and `example_data.sav` as the data file.

Example 1. Continuous DV, centered solution (x and m mean-centered, y left in original metric).

```
modprobe y=tosca /x=attitude /m=ambivalence /center=1.
```

Example 2. Continuous DV, half-standardized solution (x and m standardized, x ranging from -1SD to 1SD, y unstandardized).

```
modprobe y = tosca /x = attitude /m = ambivalence /center = 3.
```

Example 3. Similar to Example 2, but requesting conditional testing and plot at values of +/- 2 SDs of the moderator. Also, a range of -3 to +3 SDs of x is specified for the plot (use `center = 2` for this).

```
modprobe y=tosca /x=attitude /m=ambivalence /center=2 /mlow = -2 /mhigh = 2/ xlow = -3 /xhigh = 3.
```

Example 4. Continuous DV, fully-standardized solution (x, m, and y are z-standardized, values of x range from -1 to 1). Also, a filter variable (`filter1`) is added such that only cases with a 1 on the filter variable are included in analyses.

```
modprobe y=tosca /x=attitude /m=ambivalence /center=5 /addfilter = filter1.
```

Example 5. Same as Example 4, including a covariate (`self_esteem`) and requesting boxplots of y,x,m, and covariates.

```
modprobe y=tosca /x=attitude /m=ambivalence /cov = self_esteem /center=5 /boxplot = 1.
```

Example 6. Same as Example 5, excluding possible univariate (+/- 3 SDs) and multivariate outliers (using Leverage and Cook's Distance). Conventional cutoffs values are requested by specifying 9999 for multivariate statistics. A Histogram of multivariate indices is also requested.

```
modprobe y=tosca /x=attitude /m=ambivalence /cov = self_esteem /center=5 /allout = 3 /lever = 9999 /cook = 9999/histo = 1.
```

Example 7. Dichotomous DV (logistic regression), centered solution, standard plot (log-odds).

```
modprobe y=voting_decision/x=attitude/m=certainty/center=1.
```

Example 8. Dichotomous DV (logistic regression), centered solution, requesting a probability plot.

```
modprobe y=voting_decision/x=attitude/m=certainty/center=1/plotprob=1.
```