> **bmagraph varmap** — Variable-inclusion map after BMA regression

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Methods and formulas	Reference
Also see			

Description

bmagraph varmap produces a variable-inclusion map after the bmaregress command for top models ranked by the highest posterior model probability (PMP), which include predictors with posterior inclusion probability (PIP) above 0.01.

Quick start

Inclusion map for predictors included in the top 100 models ranked by highest PMP

    bmagraph varmap

Inclusion map for predictors with PIP of at least 0.2 and for top 10 models

    bmagraph varmap, top(10) pipcutoff(0.2)

Same as above, but force all model bars to have equal widths, instead of widths being proportional to the model's PMP

    bmagraph varmap, top(10) pipcutoff(0.2) equalwidths

Menu

Statistics > Bayesian model averaging > Variable-inclusion map

Syntax

>   bmagraph varmap [, *options*]

options	Description
[Main]	
top(*#*)	show top *#* models with highest PMPs; default is top(100)
pipcutoff(*#*)	do not show predictors with PIP less than *#*; default is *# = 0.01*
maxmodels(*#*)	plot results for the first *#* models; default is maxmodels(1000)
equalwidths	request that equal-width model bars are displayed; default is widths proportional to PMP
pmpnormalized	show normalized PMP
poscoefopts(*rbaropts*)	control the look of bars for positive coefficients
negcoefopts(*rbaropts*)	control the look of bars for negative coefficients
zerocoefopts(*rbaropts*)	control the look of bars for zero (not-included) coefficients
Range bar options	
rbaropts	control the look of all bars
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*
all	plot all models

all does not appear in the dialog box.

Options

[Main]

top(*#*) specifies that the top *#* models with the highest PMPs be shown. By default, the top 100 are shown. The maximum number of models plotted is controlled by option maxmodels(). Use top() in combination with maxmodels(), if you need to plot more than 1,000 models.

pipcutoff(*#*) specifies that predictors with PIPs less than *#* not be shown. The default is pipcutoff(0.01). This option is useful when there are many predictors with small PIPs.

maxmodels(*#*) specifies the maximum number of models to be plotted. The default is maxmodels(1000). When this option is specified together with option top(*#₁*), the number of plotted models is the minimum between *#* and *#₁*.

equalwidths specifies that all plotted model bars have equal widths. By default, the bar widths are proportional to PMPs. This option may not be specified together with pmpnormalized.

pmpnormalized specifies that the model bar widths correspond to the normalized PMPs of the plotted models instead of PMPs. After normalization, the plotted range of cumulative PMPs (CPMP) will be between 0 and 1. This option may not be specified together with option equalwidths.

poscoefopts(*rbaropts*), negcoefopts(*rbaropts*), and zerocoefopts(*rbaropts*) control the look of bars for the positive, negative, and zero (not-included) coefficients, respectively. *rbaropts* are any options as defined in [G-3] *barlook_options*, and barwidth(*#*), mwidth, and msize(*markersizestyle*) as defined in [G-2] **graph twoway rbar**.

⌐ Range bar options ⌐

rbaropts control the look of all bars. *rbaropts* are any options as defined in [G-3] *barlook_options*, and barwidth(#), mwidth, and msize(*markersizestyle*) as defined in [G-2] **graph twoway rbar**.

⌐ Y axis, X axis, Titles, Legend, Overall ⌐

twoway_options are any of the options documented in [G-3] *twoway_options*, excluding by(). These include options for titling the graph (see [G-3] *title_options*) and for saving the graph to disk (see [G-3] *saving_option*).

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

all specifies that all models, up to maxmodels(), be shown on the graph. This option may be useful when the number of visited models is small. This option may not be specified together with top().

Remarks and examples

stata.com

In the context of BMA, a variable-inclusion map is a graphical representation of which variables or predictors are included in individual regression models. The plotted models are ranked by the highest PMP. Each predictor (with a PIP above 0.01) within a model is represented by a colored bar: blue if a predictor is included in a model with a positive coefficient, red if a coefficient is negative, and gray if a predictor is not included in the model. This map provides a graphical summary about the included predictors and their probabilities of being included across models and the signs of their corresponding regression coefficients. We can also evaluate the complexity of models from this map based on the number of included predictors.

▷ Example 1: Variable-inclusion map after BMA regression

Recall the performance dataset (Chatterjee and Hadi 2012, sec. 3.3) analyzed in example 1 of [BMA] **bmaregress**, where the employees' satisfaction with their supervisors, rating, is modeled by six potential predictors.

Let's fit a linear BMA regression to these data by using `bmaregress` and plot a variable-inclusion map by using `bmagraph varmap`.

```
. use https://www.stata-press.com/data/r18/performance
(Data on employee satisfaction with supervisor)

. bmaregress rating complaints-advance

Enumerating models ...
Computing model probabilities ...

Bayesian model averaging                   No. of obs       =       30
Linear regression                          No. of predictors =       6
Model enumeration                                    Groups =        6
                                                     Always =        0
Priors:                                    No. of models    =       64
  Models: Beta-binomial(1, 1)                 For CPMP >= .9 =       10
   Cons.: Noninformative                   Mean model size   =    1.699
   Coef.: Zellner's g
       g: Benchmark, g = 36                Shrinkage, g/(1+g) = 0.9730
  sigma2: Noninformative                   Mean sigma2       =   52.302
```

rating	Mean	Std. dev.	Group	PIP
complaints	.7052859	.1224289	1	.99973
learning	.0603014	.1285281	3	.25249
advance	−.0167921	.073883	6	.13148
privileges	−.0074174	.0488635	2	.10998
raises	.0070789	.0670475	4	.10642
critical	.0009713	.0437848	5	.098534
Always				
_cons	14.8472	7.874219	0	1

```
Note: Coefficient posterior means and std. dev. estimated from 64 models.
Note: Default priors are used for models and parameter g.
```

From the output, the model space is fully explored through enumeration. There is a total of $2^6 = 64$ models in the full model space. For a detailed interpretation of the above output, see example 1 of [BMA] **bmaregress**.

We use the `bmagraph varmap` command to construct a detailed variable-inclusion map for the fitted BMA model.

```
. bmagraph varmap
Computing model probabilities ...
```


Variable-inclusion map

All 64 visited models shown.
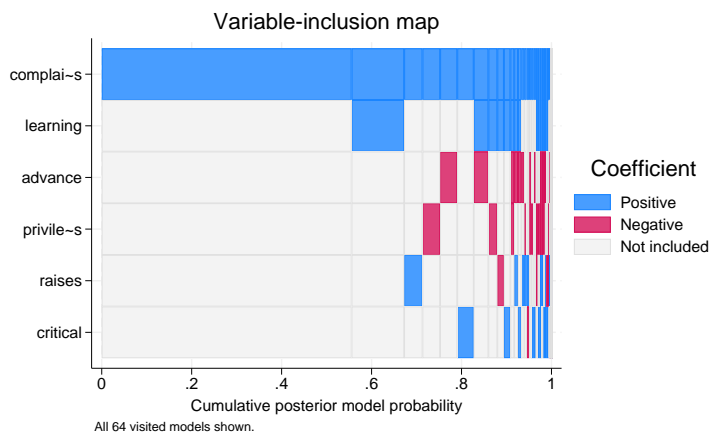
The plot displays all 64 models, which is less than the default 100. The models are ordered by their PMPs (from highest to lowest), and their CPMPs are displayed on the x axis. On the y axis, all six predictors are shown. Each pair of a model and a predictor is represented by a bar with a width that is proportional to the model's PMP. The bar is blue if a predictor is included in the model with a positive coefficient. It is red if a predictor is included in the model with a negative coefficient. It is gray if a predictor is not included in the model. In our example, because all models are included in the map, the plotted range of the CPMP x axis is between 0 and 1.

The first column corresponds to the model with the highest PMP of about 0.56 (the width of the bar), which includes only one predictor, `complaints`. This is not surprising because, from the output of `bmaregress`, predictor `complaints` has by far the highest PIP of 0.9997 of all predictors. We also see this on the map—the entire row for `complaints` has blue bars across almost all models.

The next highest-ranking model includes both `complaints` and `learning`, the predictor with the next highest PIP of 0.25. Also, it seems that the coefficients of these two predictors are always positive (blue bars) whenever they are included in a model. In other words, the two predictors appear to have a positive association with `rating`.

◁

▷ Example 2: Customizing the variable-inclusion map

The number of visited regression models can be large, and plotting all of them may become impractical. In such cases, we may use the `top(#)` option to show only the top # models with the highest PMPs. For illustration purposes, we will use `top(5)` to replicate the variable-inclusion summary table in example 1 of [BMA] **bmastats models**.

```
. bmagraph varmap, top(5)
Computing model probabilities ...
```

Variable-inclusion map

Top 5 models shown out of 64 visited.
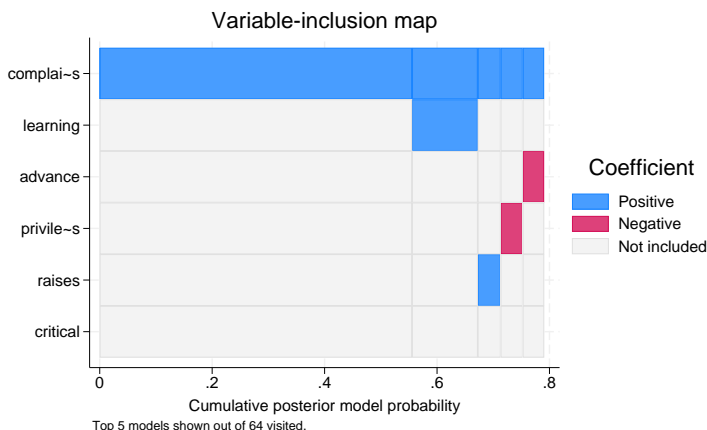
Because we are showing only 5 out of 64 models, the maximum range for the CPMP of the plotted models is less than 80%. The width of each model bar corresponds to the model's PMP value. You may specify the `pmpnormalized` option to rescale CPMP to span the range from 0 to 1.

As we mentioned earlier, variable `complaints` appears to be included in all models. However, the models with small PMPs may be represented by bars with widths too small to show the fill color. For those models, it may not be possible to determine from the map whether the predictor was included in the model. In that case, we may request that all bar widths be equal instead of being proportional to PMPs by specifying the `equalwidths` option.

```
. bmagraph varmap, equalwidths
Computing model probabilities ...
```

Variable-inclusion map
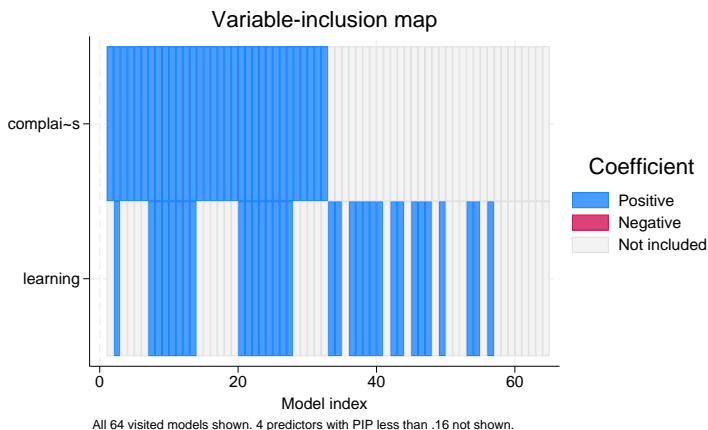

All 64 visited models shown.
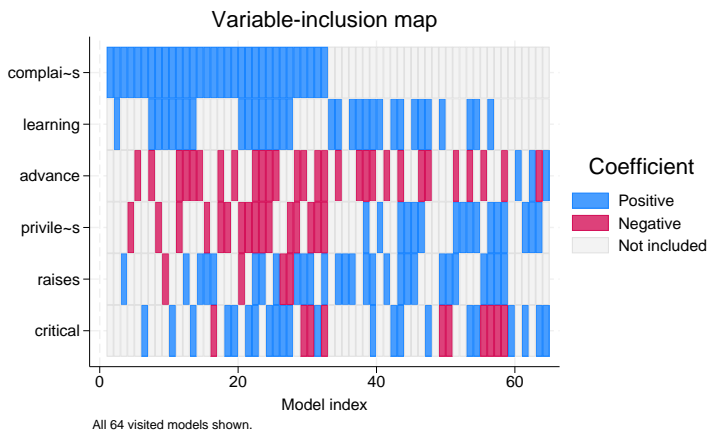
The model index (from 1 to 64) is now displayed on the x axis instead of the CPMP. It turns out that `complaints` is included only in the first 32 models with the highest PMPs with all 32 coefficients being positive.

For a BMA with many predictors and models, it may become necessary to restrict the list of the plotted predictors to those having PIP above a certain threshold to improve the readability of the

graph. Below, we use the `pipcutoff()` option to limit the plotted predictors to only those that have PIP above 0.16.

```
. bmagraph varmap, equalwidths pipcutoff(.16)
Computing model probabilities ...
```


Variable-inclusion map

All 64 visited models shown. 4 predictors with PIP less than .16 not shown.

Based on the PIP column of the `bmaregress` output in example 1, we expect four predictors (`advance`, `privileges`, `raises`, and `critical`) to be excluded from the plot because their PIP < 0.16. `bmagraph varmap` also produces a note following the command specification to inform you about the number of predictors omitted from the plot.

◁

Methods and formulas

The widths of the bars corresponding to each model are proportional to PMPs, unless the `equalwidths` option is specified. For methods and formulas of PMPs, see *Posterior model probability* in *Methods and formulas* of [BMA] **bmaregress**.

The CPMP for a model j is a cumulative sum of the first jth highest PMPs.

Reference

Chatterjee, S., and A. S. Hadi. 2012. *Regression Analysis by Example.* 5th ed. New York: Wiley.

Also see

[BMA] **bmastats models** — Model and variable-inclusion summaries after BMA regression

[BMA] **bmagraph** — Graphical summary for models and predictors after BMA regression

[BMA] **bmaregress** — Bayesian model averaging for linear regression

[BMA] **BMA postestimation** — Postestimation tools for Bayesian model averaging

[BMA] **Glossary**