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

Description

bmastats models provides summary information for the models visited by the bmaregress command, including their posterior model probabilities (PMPs), cumulative PMPs (CPMPs), ranks by PMP, and more. The variable-inclusion patterns for these models are also reported.

Quick start

Show summary statistics for the top five visited models with the highest PMPs

```
bmastats models
```

Same as above, but report CPMPs and suppress the variable-inclusion table

```
bmastats models, cumulative novartable
```

List models with CPMPs up to 0.9

```
bmastats models, cumulative(0.9)
```

Same as above, but suppress all table output for brevity, and report only the header with the number of models

```
bmastats models, cumulative(0.9) notable
```

List the top 10 models with the highest PMPs

```
bmastats models, top(10)
```

List models ranked from 10 to 15 by PMP and the median probability model (MPM)

```
bmastats models, ranks(10/15) mpm
```

Show summary statistics for models that include the two predictors x1 and x2

```
bmastats models, include(x1 x2)
```

Same as above, but display only the header with the number of reported models

```
bmastats models, include(x1 x2) notable
```

Menu

Statistics > Bayesian model averaging > Model and variable-inclusion summaries

Syntax

`bmastats models` [, *modelopts varinclopts options*]

modelopts	Description
Main	
`top(#)`	show top # models with highest PMPs; default is `top(5)`
`ranks(`*numlist*`)`	show models ranked by highest PMPs with ranks in *numlist*
`hpm`	show highest probability model (HPM) with the highest posterior probability
`mpm`	show MPM having predictors with posterior inclusion probability (PIP) ≥ 0.5
`include(`*varlist*`)`	show models that include *varlist*
`cumulative`	display CPMPs instead of PMPs for the specified models
`cumulative(#)`	display models such that the last model has CPMP of at least # (if no other options specified)
`pmpcutoff(#)`	do not show models with PMP less than #; default is # = 0
`format(%`*fmt*`)`	use numerical format %*fmt* for PMP and CPMP values
`[no]pmptable`	display or suppress model-summary table; default is `pmptable`
`all`	show all models

varlist may contain factor variables; see [U] 11.4.3 Factor variables.
`all` does not appear in the dialog box.

varinclopts	Description
Main	
`pipcutoff(#)`	do not show predictors with PIP less than #; default is `pipcutoff(.01)`
`display(x)`	indicate included predictors with an `x`; the default
`display(u)`	same as `display(x)`, except predictors not included indicated with a `u`
`sort(none)`	order of predictors as originally specified; the default
`sort(names)`	order by the names of the variables
`nolegend`	suppress table legend
`nofvlabel`	display factor-variable level values rather than value labels
`nolstretch`	do not stretch the width of the table to accommodate long variable names
`[no]vartable`	display or suppress the variable-inclusion table; default varies

options	Description
Main	
`maxmodels(#)`	display results for the first # models; default is `maxmodels(50)`
`frequency`	use frequency PMP estimates for model ranking; default is analytical PMPs (if available)
`[no]table`	display or suppress all tables; default varies

`table` and `notable` do not appear in the dialog box.

`collect` is allowed; see [U] 11.1.10 Prefix commands.

Options

Main

`top(#)` specifies that the top # models with highest PMPs be shown. By default, the top five are shown. `top(#)` is equivalent to `ranks(1/#)`.

`ranks(`*numlist*`)` specifies that models with ranks in *numlist* be included. Models are ranked by highest PMPs.

`hpm` specifies that the model with the highest PMP be shown. It is known as the HPM.

`mpm` specifies that the MPM be shown. MPM is the model in which all predictors have PIPs greater or equal to 0.5. The MPM may not always exist.

`include(`*varlist*`)` specifies that models that include all of *varlist* be shown. *varlist* may contain factor variables; see **[U] 11.4.3 Factor variables**.

Options `top()`, `ranks()`, `hpm`, `mpm`, and `include()`, when specified together, provide results for all models that meet any of the specified criteria, that is, for the union of the models.

`cumulative` and `cumulative(#)` specify that CPMPs be reported for the models instead of the default PMPs. In the absence of other options that identify models such as `ranks()` or `include()`, `cumulative(#)` displays all models for which CPMPs are less than or equal to #. If there is no model with CPMP exactly equal to #, the model with the next highest CPMP is reported as the last model. If `cumulative(#)` is specified with other options, the results are reported for the intersection of the models provided by `cumulative()` and models provided by other options, regardless of whether the specified CPMP cutoff value is reached.

`pmpcutoff(#)` specifies that models with PMP less than # not be shown. The default is `pmpcutoff(0)`. This option is useful when there are many models with small PMPs.

`format(%`*fmt*`)` specifies how numeric PMP and CPMP values are to be formatted. The default is `format(%9.0g)`.

`pmptable` and `nopmptable` specify whether to display the model-summary table. The default is `pmptable`, which displays the table for the first 50 models. You can change the maximum number of displayed models by specifying the `maxmodels()` option.

`pipcutoff(#)` specifies that predictors with PIPs less than # not be shown in the variable-inclusion table. The default is `pipcutoff(0.01)`. This option is useful when there are many predictors with small PIPs.

`display(`*displayspec*`)` specifies what to display in the variable-inclusion table. The default is `display(x)`.

Blank cells in the table indicate that the corresponding variable or predictor was not included in the model.

For some predictors without estimated coefficients, a code that indicates the reason for omission is reported in the table.

Empty levels of factors and interactions are coded with the letter `e`.

Base levels of factors and interactions are coded with the letter `b`.

Variables omitted because of collinearity are coded with the letter `o`.

`display(x)` displays an `x` in the table when the variable or predictor has been included in the displayed model.

`display(u)` is the same as `display(x)`, except that when a variable or predictor was not included in the model, `u` (for unavailable) is displayed instead of a blank cell.

`sort(`*sortspec*`)` specifies that the rows of the variable-inclusion table be ordered by specification given by *sortspec*.

`sort(none)` specifies that the rows of the table be ordered by the order the predictors were specified in the model specification. This is the default.

`sort(names)` orders rows alphabetically by the variable names of the predictors. For factor variables, main effects and nonfactor variables are displayed first in alphabetical order. Then, all two-way interactions are displayed in alphabetical order, then, all three-way interactions, and so on.

`nolegend` specifies that the legend at the bottom of the variable-inclusion table not be displayed. By default, it is shown.

`nofvlabel` displays factor-variable level numerical values rather than attached value labels. This option overrides the `fvlabel` setting. See [R] **set showbaselevels**.

`nolstretch` specifies that the width of the variable-inclusion table not be automatically widened to accommodate long variable names. When `nolstretch` is specified, names are abbreviated to make the table width no more than 79 characters. The default, `lstretch`, is to automatically widen the table up to the width of the Results window. To change the default, use `set lstretch off`.

`vartable` and `novartable` specify whether to display the variable-inclusion table. By default, the table is displayed when the number of the reported models does not exceed 12. You can specify `vartable` to display models beyond 12. The maximum number of models that will be displayed is 50, but you can change this by specifying the `maxmodels()` option.

`maxmodels(`*#*`)` specifies the maximum number of models to be displayed in model-summary and variable-inclusion tables. The default is to display the first 50 models.

`frequency` specifies that frequency estimates of PMPs based on a Markov chain Monte Carlo (MCMC) sample be used for model ranking when sampling is used by `bmaregress`. By default, analytical PMPs are used whenever they are available. Option `frequency` is not relevant with model enumeration, because frequency PMPs are not available. This option, however, is implied in the case of a random g parameter of Zellner's g-prior for regression coefficients, because analytical PMPs are not available in that case. See example 2 for details.

The following options are available with `bmastats models` but are not shown in the dialog box:

`all` specifies that all models, up to `maxmodels()`, be displayed. This option may be useful when the number of visited models is small.

`table` and `notable` display or suppress both the model-summary table and the variance-inclusion table. `table` implies `pmptable` and `vartable`. And `notable` implies `nopmptable` and `novartable`.

Remarks and examples

stata.com

In the Bayesian model averaging (BMA) framework, inference includes exploring PMPs. Models with high PMPs and the predictors they include are often of interest. Clarke (2003), among others, comments that BMA puts the most weight on the model closest to the data-generating model whether the latter is in the explored model space or not. Thus, the models with the highest PMPs, HPMs, are often of interest.

Another model of potential interest is the MPM (Barbieri and Berger 2004). This model can be viewed as a model that includes only influential predictors, that is, predictors that are more likely than not to be included in the model. Such predictors have PIPs of at least 0.5. The MPM may not always exist.

By default, `bmastats models` reports the top five models with the highest PMPs and the predictors included in each of these models. You can use the `top()` option to display more top models and `ranks()` to display models with specific ranks. You can use options `hpm` and `mpm` to display the respective HPM and MPM. And you can use the `include(`*varlist*`)` option to display models that contain the specified predictors in *varlist* and their respective PMP rankings. When you combine all of these options, the union of the models is displayed.

The command reports PMPs by default, but you can specify the `cumulative` option to report CPMPs instead. The `cumulative(#)` option is useful if you wish to see the models with CPMPs up to #. This option (as well as the `include()` option), however, may report many models, especially for high # values. You may want to combine it with the `notable` option to see only the number of models to be reported first. You can also use the `pmpcutoff(#)` option to limit the models to those with PMPs of at least #.

`bmastats models` reports two tables for the models: one with the PMPs and the other one with variable-inclusion patterns. The variable-inclusion table can be rather lengthy depending on the number of included predictors. By default, only predictors with PIPs above 0.01 are displayed, but you can change this with the `pipcutoff()` option. Also, this table is suppressed when the number of reported models exceeds 12. You can use the `vartable` option to display more models. The maximum number of reported models for both tables is 50, but you can change this by specifying `maxmodels()`.

See example 1 for various uses of `bmastats models`'s options.

`bmastats models` reports analytical and frequency PMPs, whenever they are available. Analytical PMPs are computed using analytical formulas. They are reported only with a fixed *g* because analytical formulas are not available with a random *g*. Frequency PMPs are computed from an MCMC sample of models, which is available only when sampling is used by `bmaregress`. That is, they are not available with model enumeration. Analytical PMPs are used for model ranking whenever they are available. See example 2.

▷ Example 1: Tour of the bmastats models command

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

Let's fit a linear BMA regression to these data using `bmaregress` and explore various model summaries by using `bmastats models`.

```
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.

. bmastats models

Computing model probabilities ...

Model summary             Number of models:
                              Visited = 64
                             Reported =  5

--------------------------------------------
             |  Analytical PMP    Model size
-------------+------------------------------
Rank         |
           1 |           .5556             1
           2 |           .1169             2
           3 |          .04072             2
           4 |          .03932             2
           5 |          .03804             2
--------------------------------------------

Variable-inclusion summary

------------------------------------------------------------------
             |    Rank       Rank       Rank       Rank       Rank
             |     1          2          3          4          5
-------------+----------------------------------------------------
  complaints |     x          x          x          x          x
    learning |                x
      raises |                           x
  privileges |                                      x
     advance |                                                 x
------------------------------------------------------------------
Legend:
  x - estimated
```

By default, `bmastats models` reports the top five models ranked by PMPs and their respective variable-inclusion summary. We see that the top model, also known as HPM, has a PMP of 0.56 and includes only one predictor, `complaints`. This is not surprising because, from the output of `bmaregress`, predictor `complaints` has by far the highest PIP, 0.9997, of all predictors. The next highest-ranking model, with a PMP of 0.12, includes both `complaints` and `learning`, the predictor with the next highest PIP, 0.25. The rest of the predictors have similar probabilities, which are less than 0.15.

We can display more than five top models by specifying the `top()` option. Here, to demonstrate, we list only the top six models to keep the output short.

```
. bmastats models, top(6)

Computing model probabilities ...

Model summary              Number of models:
                                     Visited = 64
                                    Reported =  6
```

	Analytical PMP	Model size
Rank		
1	.5556	1
2	.1169	2
3	.04072	2
4	.03932	2
5	.03804	2
6	.03654	2

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
complaints	x	x	x	x	x	x
learning		x				
raises			x			
privileges				x		
advance					x	
critical						x

```
Legend:
  x - estimated
```

The above is equivalent to `ranks(1/6)`, or you can specify any list of ranks if desired.

```
. bmastats models, ranks(1/6)
Computing model probabilities ...
Model summary                Number of models:
                                 Visited = 64
                                Reported =  6
```

	Analytical PMP	Model size
Rank		
1	.5556	1
2	.1169	2
3	.04072	2
4	.03932	2
5	.03804	2
6	.03654	2

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
complaints	x	x	x	x	x	x
learning		x				
raises			x			
privileges				x		
advance					x	
critical						x

```
Legend:
  x - estimated
```

We can also use options `hpm` and `mpm` to display the respective HPM and MPM.

```
. bmastats models, hpm mpm
Computing model probabilities ...
Model summary                Number of models:
                                 Visited = 64
                                Reported =  1
```

	Analytical PMP	Model size
Rank		
(HPM) 1	.5556	1

```
Note: HPM and MPM are the same model.
```

Variable-inclusion summary

	(HPM) 1
complaints	x

```
Legend:
  x - estimated
```

In our example, HPM and MPM are the same.

Instead of PMPs, we can use the `cumulative` option to display CPMPs.

```
. bmastats models, cumulative
Computing model probabilities ...
Model summary              Number of models:
                                    Visited = 64
                                   Reported =  5

─────────────┬─────────────────────────────────
             │  Analytical CPMP     Model size
─────────────┼─────────────────────────────────
Rank         │
           1 │            .5556              1
           2 │            .6724              2
           3 │            .7132              2
           4 │            .7525              2
           5 │            .7905              2
─────────────┴─────────────────────────────────

Variable-inclusion summary
─────────────┬───────────────────────────────────────────────────────────────
             │    Rank       Rank       Rank       Rank       Rank
             │      1          2          3          4          5
─────────────┼───────────────────────────────────────────────────────────────
  complaints │      x          x          x          x          x
    learning │                 x
      raises │                            x
  privileges │                                       x
     advance │                                                  x
─────────────┴───────────────────────────────────────────────────────────────
Legend:
  x - estimated
```

You might also want to see the models that contribute to a specific, typically larger, CPMP value. For instance, let's use 0.9 for the CPMP cutoff here.

```
. bmastats models, cumulative(0.9)
Computing model probabilities ...
Model summary                Number of models:
                                       Visited = 64
                                      Reported = 10
```

	Analytical CPMP	Model size
Rank		
1	.5556	1
2	.6724	2
3	.7132	2
4	.7525	2
5	.7905	2
6	.8271	2
7	.8598	3
8	.8796	3
9	.8942	3
10	.9086	3

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
complaints	x	x	x	x	x	x
learning		x				
raises			x			
privileges				x		
advance					x	
critical						x

Legend:
x - estimated

	Rank 7	Rank 8	Rank 9	Rank 10
complaints	x	x	x	x
learning	x	x	x	x
raises			x	
privileges		x		
advance	x			
critical				x

Legend:
x - estimated

There are 10 models that contribute to the CPMP of at least 0.9. This is the same number as reported in the `bmaregress` header output under `For CPMP >= 0.9`. This number is useful to determine whether there are only a few high-probability models that are consistent with your data or there are many different models that are plausible.

Beware that `cumulative()` may report many models, especially for high CPMP values. In that case, to keep the table output manageable, `bmastats models` will display only the first 50 regardless of the specified CPMP value. But you can use the `maxmodels()` option to display more. You might also consider using the `notable` option to suppress the table output altogether, if you just want to see the number of models corresponding to a specific CPMP value. We demonstrate this later in combination with the `include()` option, which also may lead to many models being displayed.

In fact, for brevity, the variance-inclusion table is suppressed when the number of reported models exceeds 12, as we see in the output.

```
. bmastats models, cumulative(0.95)
Computing model probabilities ...
Model summary              Number of models:
                                    Visited = 64
                                   Reported = 16
```

	Analytical CPMP	Model size
Rank		
1	.5556	1
2	.6724	2
3	.7132	2
4	.7525	2
5	.7905	2
6	.8271	2
7	.8598	3
8	.8796	3
9	.8942	3
10	.9086	3
11	.9172	4
12	.9255	4
13	.9333	4
14	.9396	3
15	.9452	3
16	.9503	3

```
Note: Use option vartable to display variable-inclusion table for more than
      12 models.
```

But you can specify the `vartable` option to see those models.

```
. bmastats models, cumulative(0.95) vartable
Computing model probabilities ...
Model summary              Number of models:
                                      Visited = 64
                                     Reported = 16
```

	Analytical CPMP	Model size
Rank		
1	.5556	1
2	.6724	2
3	.7132	2
4	.7525	2
5	.7905	2
6	.8271	2
7	.8598	3
8	.8796	3
9	.8942	3
10	.9086	3
11	.9172	4
12	.9255	4
13	.9333	4
14	.9396	3
15	.9452	3
16	.9503	3

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
complaints	x	x	x	x	x	x
learning		x				
raises			x			
privileges				x		
advance					x	
critical						x

Legend:
x - estimated

	Rank 7	Rank 8	Rank 9	Rank 10	Rank 11	Rank 12
complaints	x	x	x	x	x	x
learning	x	x	x	x	x	x
raises			x			x
privileges		x			x	
advance	x				x	x
critical				x		

Legend:
x - estimated

	Rank 13	Rank 14	Rank 15	Rank 16
complaints	x	x	x	x
learning	x			
raises		x	x	x
privileges			x	
advance	x	x		
critical	x			x

```
Legend:
  x - estimated
```

We can explore the models that contain specific predictors by using the `include()` option. We use the `complaints` variable. We also specify the `notable` option to see only the number of models to be reported first.

```
. bmastats models, include(complaints) notable
Computing model probabilities ...
Model summary                Number of models:
                                      Visited = 64
                                     Reported = 32
```

There are 32 different models that include `complaints`. We probably do not want to display all 32 models, because many of them will have a low probability. We can use the `pmpcutoff()` option to display only models with higher probabilities.

```
. bmastats models, include(complaints) pmpcutoff(0.01)
Computing model probabilities ...
Model summary             Number of models:
                                   Visited = 64
                                  Reported = 10

─────────────┬─────────────────────────────
             │   Analytical PMP   Model size
─────────────┼─────────────────────────────
Rank         │
           1 │            .5556            1
           2 │            .1169            2
           3 │           .04072            2
           4 │           .03932            2
           5 │           .03804            2
           6 │           .03654            2
           7 │           .03272            3
           8 │           .01985            3
           9 │           .01459            3
          10 │           .01441            3
─────────────┴─────────────────────────────
Note: 22 models with PMP less than .01 not shown.
Variable-inclusion summary

─────────────┬───────────────────────────────────────────────────────────
             │    Rank      Rank      Rank      Rank      Rank      Rank
             │       1         2         3         4         5         6
─────────────┼───────────────────────────────────────────────────────────
  complaints │       x         x         x         x         x         x
    learning │                 x
      raises │                           x
  privileges │                                     x
     advance │                                               x
    critical │                                                         x
─────────────┴───────────────────────────────────────────────────────────
Legend:
  x - estimated

─────────────┬───────────────────────────────────────
             │    Rank      Rank      Rank      Rank
             │       7         8         9        10
─────────────┼───────────────────────────────────────
  complaints │       x         x         x         x
    learning │       x         x         x         x
      raises │                           x
  privileges │                 x
     advance │       x
    critical │                                     x
─────────────┴───────────────────────────────────────
Legend:
  x - estimated
```

We could further explore a few higher-probability models in more detail by specifying their ranks in the `ranks()` option and looking at their variable-inclusion summary. The top two models above are the same ones we already explored earlier: the one containing `complaints` and the other one containing `complaints` and `learning`.

◁

▷ Example 2: Analytical and frequency PMPs

In this example, we explore the analytical and frequency PMPs as reported by `bmastats models`.

Let's revisit example 1, where model enumeration was used for estimation. In what follows, we will use the `novartable` option with `bmastats models` to focus only on the reported PMPs.

```
. 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.

. bmastats models, novartable

Computing model probabilities ...

Model summary              Number of models:
                                    Visited = 64
                                   Reported =  5

──────────────────────────────────────────────
             │   Analytical PMP     Model size
─────────────┼────────────────────────────────
Rank         │
           1 │            .5556              1
           2 │            .1169              2
           3 │           .04072              2
           4 │           .03932              2
           5 │           .03804              2
──────────────────────────────────────────────
```

With a small number of predictors, `bmaregress` explores a full space of all possible models, so no sampling is performed. In this case, `bmastats models` reports only analytical PMPs.

With many predictors, a full model enumeration may not be feasible, and thus a sampling algorithm is used to explore the model space. To demonstrate, let's use sampling instead of the default model enumeration in our example by specifying the `sampling` option with `bmaregress`.

```
. bmaregress rating complaints-advance, sampling rseed(18)
Burn-in ...
Simulation ...
Computing model probabilities ...

Bayesian model averaging                          No. of obs         =     30
Linear regression                                 No. of predictors  =      6
MC3 sampling                                                 Groups =      6
                                                             Always =      0
                                                  No. of models      =     32
                                                      For CPMP >= .9 =     10
Priors:                                           Mean model size    =  1.699
  Models: Beta-binomial(1, 1)                     Burn-in            =  2,500
   Cons.: Noninformative                          MCMC sample size   = 10,000
   Coef.: Zellner's g                             Acceptance rate    = 0.2417
       g: Benchmark, g = 36                       Shrinkage, g/(1+g) = 0.9730
  sigma2: Noninformative                          Mean sigma2        = 52.292

Sampling correlation = 0.9990

-----------------------------------------------------------------------------
      rating |      Mean   Std. dev.                         Group        PIP
-------------+---------------------------------------------------------------
  complaints |   .705479   .1218881                              1          1
    learning |  .0601919   .1282869                              3     .25234
     advance | -.0167514   .0737415                              6     .13141
  privileges | -.0074265    .048844                              2     .10996
      raises |  .0069949   .0666406                              4     .10629
    critical |  .0009699   .0437742                              5    .098526
-------------+---------------------------------------------------------------
Always       |
       _cons |  14.84478   7.871046                              0          1
-----------------------------------------------------------------------------
Note: Coefficient posterior means and std. dev. estimated from 32 models.
Note: Default priors are used for models and parameter g.

. bmastats models, novartable

Computing model probabilities ...

Model summary                                Number of models:
                                                       Visited = 32
                                                      Reported =  5

----------------------------------------------------------------
             |  Analytical PMP   Frequency PMP      Model size
-------------+--------------------------------------------------
Rank         |
           1 |           .5557           .5167               1
           2 |           .1169           .1248               2
           3 |          .04073           .0427               2
           4 |          .03933           .0392               2
           5 |          .03805           .0492               2
----------------------------------------------------------------
Note: Using analytical PMP for model ranking.
```

Analytical and frequency PMPs are available in this case, and `bmastats models` reports both. The analytical and frequency estimates are similar, which should be the case for a converged model. `bmaregress` reported a sampling correlation of 0.9990, which is a strong indication of convergence. When both types of PMP estimates are available, `bmastats models` always uses analytical PMPs for model ranking, but you can use the `frequency` option to rank models by frequency PMPs.

Both BMA models above assumed a fixed parameter g that controls the shrinkage of coefficients toward zero. We can instead specify a prior distribution, a hyperprior, for g. For instance, we use a robust hyperprior below.

```
. bmaregress rating complaints-advance, gprior(robust) rseed(18)
Burn-in ...
Simulation ...
Computing model probabilities ...
Bayesian model averaging                          No. of obs         =     30
Linear regression                                 No. of predictors =      6
MC3 and adaptive MH sampling                                  Groups =      6
                                                              Always =      0
                                                  No. of models      =     34
                                                     For CPMP >= .9 =      12
Priors:                                           Mean model size    =  1.734
  Models: Beta-binomial(1, 1)                     Burn-in            =  2,500
   Cons.: Noninformative                          MCMC sample size   = 10,000
   Coef.: Zellner's g                             Acceptance rate    = 0.4232
       g: Robust
  sigma2: Noninformative                          Mean sigma2        = 53.095
Sampling correlation = 0.9994

       rating        Mean   Std. dev.                          Group      PIP

   complaints    .7000463   .1273543                               1    .9998
     learning    .0594904   .1286095                               3      .25
      advance   -.0192712   .0797935                               6    .1503
       raises    .0079416   .0727859                               4    .1201
   privileges   -.0072591   .0487009                               2    .1069
     critical    .0014397   .0466476                               5    .1067

Always
        _cons    15.24911   7.988166                               0        1

Note: Coefficient posterior means and std. dev. estimated from 34 models.
Note: Default prior is used for models.

                                                           Equal-tailed
                 Mean   Std. dev.      MCSE     Median  [95% cred. interval]

            g  152.668   1968.132   43.5265   33.81024   8.205076   610.6026
    Shrinkage .9656427   .0276071   .001234   .9712728   .8913639   .9983649

. bmastats model, novartable
Computing model probabilities ...
Model summary             Number of models:
                                   Visited = 34
                                  Reported =  5

                   Frequency PMP     Model size

Rank
            1            .5708              1
            2             .102              2
            3            .0403              2
            4              .04              2
            5            .0305              2
```

Only frequency PMPs are reported in this case because analytical formulas for PMPs are not available for a random g.

◁

Stored results

`bmastats models` stores the following in `r()`:

Scalars	
`r(k_models)`	number of models
`r(k_models_rpt)`	number of reported models
Macros	
`r(pmptype)`	`analytical` or `frequency`
Matrices	
`r(summary)`	model summary matrix
`r(rank)`	model ranks
`r(varinclmat)`	variable-inclusion matrix

Methods and formulas

For methods and formulas of PMPs and PIPs, see *Posterior model probability* and *Posterior inclusion probability* in *Methods and formulas* of [BMA] **bmaregress**.

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

The HPM is the model with the highest PMP.

The MPM is the model with predictors that have PIPs greater than or equal to 0.5 (Barbieri and Berger 2004). That is, if $X_1, X_2, \ldots, X_p$ are predictors used with `bmaregress`, then X_k is in the MPM if and only if $\text{PIP}(X_k) \geq 0.5$. If there are no such predictors, then MPM does not exist.

References

Barbieri, M. M., and J. O. Berger. 2004. Optimal predictive model selection. *Annals of Statistics* 32: 870–897. https://doi.org/10.1214/009053604000000238.

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

Clarke, B. 2003. Comparing Bayes model averaging and stacking when model approximation error cannot be ignored. *Journal of Machine Learning Research* 4: 683–712.

Also see

[BMA] **bmagraph pmp** — Model-probability plots after BMA regression

[BMA] **bmagraph varmap** — Variable-inclusion map after BMA regression

[BMA] **bmastats** — 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**