mwp-468

Intro - ベイズモデル平均化機能の紹介　【 評価版 】

本 whitepaper ではベイズモデル平均化 (BMA: Bayesian model averaging) 機能について紹介します。BMA 関連のコマンド一覧については [BMA] **BMA commands** (*mwp-469*) を、ベイズ分析全般については [BAYES] **Intro** (BY01) を参照ください。

1. 背景
2. 機能概要
3. ベイズモデル平均化 (BMA)
4. BMA の概念
5. BMA の用途
6. 頻度論的アプローチとの比較
7. BMA の計算法
8. BMA の用例
   - Example 1
   - Example 2
   - Example 3
9. 文献レビュー

1. 背景

モデル平均化 (model averaging) というのはモデルの不確定性に対処するための統計的手法です。この手法では単一のモデルに依存するのではなく、複数のもっともらしいモデルによる分析結果を平均化するというアプローチを取ります。ベイズモデル平均化 (BMA: Bayesian model averaging) の場合、モデルの“もっともらしさ”は事後モデル確率 (posterior model probability) — 普遍的なベイズの定理 (Bayes theorem) に基づき誘導される値 — によって表現されます。

モデル平均化という手法はモデルパラメータの推定や予測に際して過度に楽観的な結論が導かれるのを回避すべく、モデルの不確定性 (model uncertainty) を考慮する目的で用いられます。特にもっともらしいモデルが複数考えられる場合に有効です。最終的に 1 つのモデルに絞り込むことが目標であったにしても、モデル平均化というプロセスを踏むことによって種々の有用な情報が抽出できることがあります。

2. 機能概要

統計学においてはモデルという概念が中心的存在となります。何らかのデータ生成モデル (DGM: data-generating model) が存在し、その特性は観測されたデータから推論し得るという考え方が統計的推論の基盤となっているためです。このため統計的な分析を行う際には、直面する問題に対し適切なモデルを選択することが最初の、かつ最も重要なステップとなります。問題によっては DGM に関し明確な理論的、経験的な裏付けが存在するケースもあるでしょう。しかし信頼のおける単一モデルを選択することが難しいケースも存在します。そのような場合にはモデル選択という過程における不確定性を考慮に入れた節理ある手法が求められます。現実には DGM の特定の性質や量に関心が集中してしまうことがよくあります。古典的な推論手法では 1 つのモデルを選択した上で、この量を選択されたモデルに条件付ける形で観測データから推定するわけです。単一モデルをベースとするこのアプローチの 1 つの欠点は、推定結果に対しデータによってサポートされるレベル以上の精度をアサインしてしまいがちである点にあります。

モデル平均化のアプローチはこれとは考え方を異にするものです。1 つのモデルを選択するのではなく、いくつかのモデル候補について検討を行います。関心対象の量は個々のモデル推定値の平均という形で推定されることになります。平均化に際してはそれぞれのモデルの確からしさが重みとして用いられます。このようにしてモデル選択に伴う不確定性への対処が図られるわけです。

3. ベイズモデル平均化 (BMA)

評価版では割愛しています。

4. BMA の概念

評価版では割愛しています。

5. BMA の用途

評価版では割愛しています。

6. 頻度論的アプローチとの比較

評価版では割愛しています。

7. BMA の計算法

評価版では割愛しています。

8. BMA の用例

ここでは人為的にシミュレートされた Example データセット bmaintro.dta を用いて BMA の用例を紹介します。

```
. use https://www.stata-press.com/data/r19/bmaintro.dta [*1]
(Simulated data for BMA example)
```

このデータセット中には $n = 200$ 件の観測データが記録されています。また予測変数の数は $p = 10$ です。x1 から x10 までの予測変数は標準正規分布に従う形で独立に生成されています。一方、アウトカム変数 y は次の回帰モデル (DGM) を用いて生成されています。

$$\mathrm{y} = 0.5 + 1.2 \times \mathrm{x2} + 5 \times \mathrm{x10} + \epsilon$$

ただし $\epsilon \sim N(0, 1)$ であるとします。

```
. summarize
```

```
. summarize

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
           y |        200    .9944997    4.925052    -13.332   13.06587
          x1 |        200   -.0187403    .9908957  -3.217909   2.606215
          x2 |        200   -.0159491    1.098724  -2.999594   2.566395
          x3 |        200    .080607    1.007036  -3.016552   3.020441
          x4 |        200    .0324701    1.004683  -2.410378   2.391406
-------------+---------------------------------------------------------
          x5 |        200   -.0821737    .9866885  -2.543018   2.133524
          x6 |        200    .0232265    1.006167  -2.567606   3.840835
          x7 |        200   -.1121034    .9450883  -3.213471   1.885638
          x8 |        200   -.0668903    .9713769  -2.871328   2.808912
          x9 |        200   -.1629013    .9550258  -2.647837   2.472586
-------------+---------------------------------------------------------
         x10 |        200    .083902    .8905923  -2.660675   2.275681
```

[*1] メニュー操作：File ▷ Example Datasets ▷ Stata 19 manual datasets と操作、Bayesian Model Averaging Reference Manual [BMA] の Intro の項よりダウンロードする。

▷ Example 1: BMA **線形回帰**

最初に y を従属変数、x1–x10 を独立変数とする形で標準的な線形回帰を実行してみます。

```
. regress y x1-x10
```

```
. regress y x1-x10

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(10, 189)      =    396.30
       Model |  4607.24837        10  460.724837   Prob > F        =    0.0000
    Residual |  219.723235       189   1.1625568   R-squared       =    0.9545
-------------+----------------------------------   Adj R-squared   =    0.9521
       Total |   4826.9716       199  24.2561387   Root MSE        =    1.0782

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          x1 |   .0753537   .0781737     0.96   0.336    -.0788513    .2295587
          x2 |    1.18854   .0716658    16.58   0.000     1.047172    1.329907
          x3 |  -.1871012   .0789484    -2.37   0.019    -.3428344   -.0313679
          x4 |  -.0459335   .0785503    -0.58   0.559    -.2008813    .1090144
          x5 |   .0343498   .0793095     0.43   0.665    -.1220956    .1907953
          x6 |  -.0149194   .0767357    -0.19   0.846    -.1662879     .136449
          x7 |    .007174   .0831239     0.09   0.931    -.1567958    .1711437
          x8 |  -.0384917   .0810626    -0.47   0.635    -.1983953    .1214119
          x9 |   .0968948   .0817218     1.19   0.237    -.0643093    .2580989
         x10 |    5.13251   .0877447    58.49   0.000     4.959426    5.305595
       _cons |    .617996   .0791152     7.81   0.000     .4619337    .7740582
------------------------------------------------------------------------------
```

regress は真の予測変数である x2 と x10 を共に統計的に有意と判定しています。x2 の係数値は 1.19、標準誤差は 0.072、95% CI は $[1.05, 1.33]$ と推定されていますが、真の値が 1.2 であることを考慮するなら良好な推定結果と言えるでしょう。また x10 については係数値が 5.13、標準誤差が 0.088、95% CI が $[4.96, 5.31]$ と推定されており、これも真の値である 5 と良く一致しています。一方で regress は x3 について 0.019 という p 値をレポートしてきています。係数値は -0.19、95% CI は $[-0.34, -0.03]$ と推定されているわけですが、これは DGM と不整合な結果であると言えます。regress からの出力について言えば、レポートされている p 値から予測変数の重要性を推し量りたくなるわけですが、p 値をそのように解釈するのは誤りです。

次に `bmaregress` コマンドを使って BMA を実行してみます。

- Statistics ▷ Bayesian model averaging ▷ Linear regression と操作
- Model タブ: Dependent variable: `y`
  In-out predictors: `x1-x10`

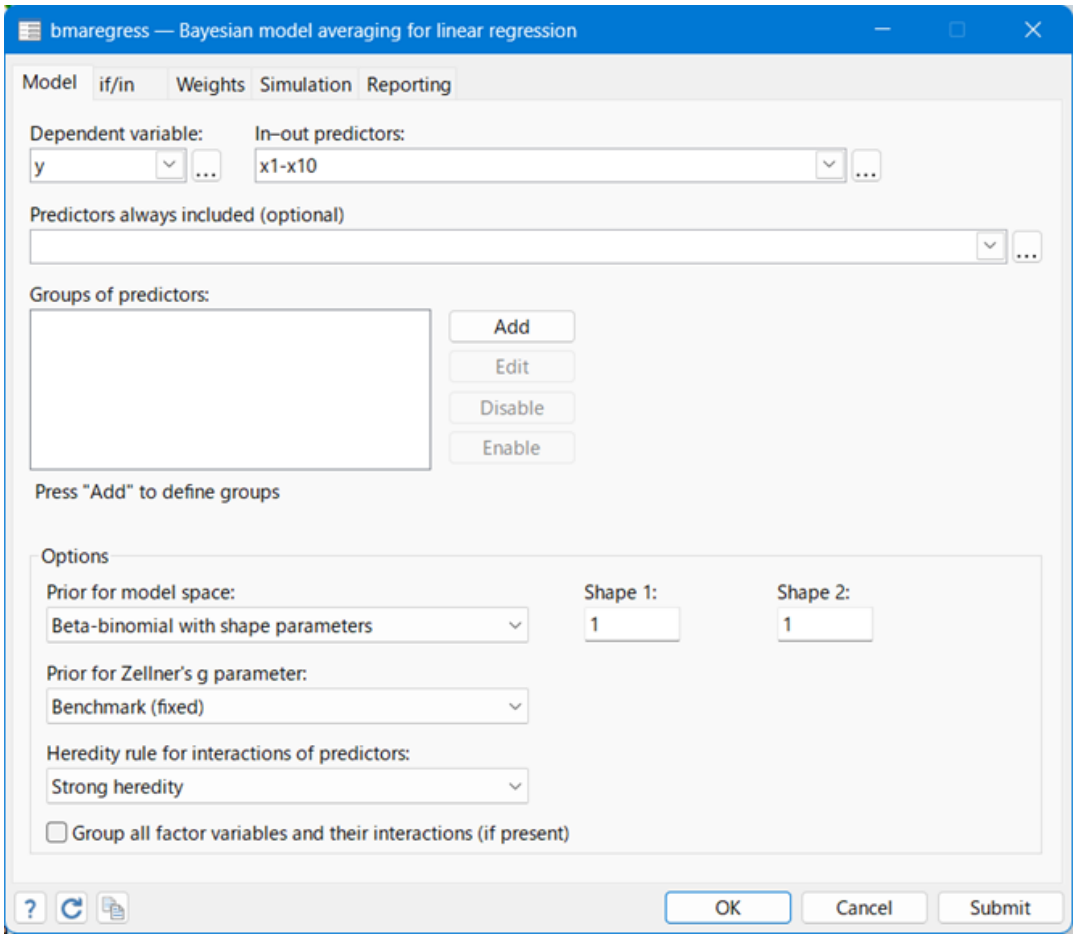

図 1 `bmaregress` ダイアログ - Model タブ

```
. bmaregress y x1-x10

Enumerating models ...
Computing model probabilities ...

Bayesian model averaging                        No. of obs         =    200
Linear regression                               No. of predictors  =     10
Model enumeration                                           Groups =     10
                                                            Always =      0
Priors:                                         No. of models      =  1,024
  Models: Beta-binomial(1, 1)                       For CPMP >= .9 =      9
   Cons.: Noninformative                        Mean model size    =  2.479
   Coef.: Zellner's g
       g: Benchmark, g = 200                    Shrinkage, g/(1+g) = 0.9950
  sigma2: Noninformative                        Mean sigma2        =  1.272

------------------------------------------------------------------------------
           y |      Mean   Std. dev.                      Group        PIP
-------------+----------------------------------------------------------------
          x2 |  1.198105   .0733478                          2          1
         x10 |   5.08343   .0900953                         10          1
          x3 | -.0352493   .0773309                          3     .21123
          x9 |   .004321   .0265725                          9    .051516
          x1 |  .0033937   .0232163                          1    .046909
          x4 | -.0020407   .0188504                          4    .039267
          x5 |  .0005972   .0152443                          5    .033015
          x8 | -.0005639   .0153214                          8    .032742
          x7 | -8.23e-06    .015497                          7    .032386
          x6 | -.0003648   .0143983                          6    .032361
-------------+----------------------------------------------------------------
Always       |
       _cons |  .5907923   .0804774                          0          1
------------------------------------------------------------------------------
Note: Coefficient posterior means and std. dev. estimated from 1,024 models.
Note: Default priors are used for models and parameter g.
```

ここでは最も関連のある情報についてのみ記述します。bmaregress からの出力の詳細については [BMA] **bmaregress** (*mwp-470*) の Example 1 を参照ください。

bmaregress はデフォルト設定の場合、予測変数の数が 10 であることに基づき、$2^{10} = 1,024$ 個のモデルについて検討を加えます。テーブル上、x2 と x10 に対する PIP が 1 と推定されている点に注意してください。一方、他の予測変数に対する PIP 値は x3 を除き、すべて 10% より小さな値となっています。そしてこの PIP 値は予測変数の重要度を示す指標として解釈することができます。PIP は 1,024 個のモデル空間を通じて、ある予測変数がモデル中に含まれる確率を表しています。従って x3 に対する 0.2 という PIP 値は 1 と比べるとかなり小さな値であるので、x3 という予測変数は余り重要とは言えないと結論付けることができるわけです。また x3 に対する BMA 係数値（事後平均）は −0.035 と推定されていますが、これは regress のときの −0.19 と比べ、より 0 に近い値となっています。

x2 と x10 の係数値についての BMA 推定値である 1.2 と 5.1 という値は真の値である 1.2 と 5 に近いものとなっています。一方、事後標準偏差推定値は 0.073 と 0.090 とレポートされていますが、これらは regress からの推定値に比べわずかに大きな値となっています。これはどの予測変数をモデル中に含めるかに関する不

確定性を BMA が考慮に入れていることから期待されることではあります。デフォルトの場合、`bmaregress` からは確信区間 (credible intervals) に関する情報が出力されてきません。これは計算コストを考慮しての対応です。もちろんそれを出力させるオプションも用意されています（`[BMA] bmaregress` (*mwp-470*) の Example 5 を参照)。なお、BMA 回帰係数の解釈に絡む留意点については **bmaregress** (*mwp-470*) を参照ください。

BMA はモデルの“選択”を行うわけではありませんが、平均化された結果に対しより寄与率の高いモデルのいくつかを特定しています。それは `bmaregress` 出力中の PIP の値から推測することができますが、`bmastats models` を使うとより鮮明な結果を得ることができます。

- Statistics ▷ Bayesian model averaging ▷ Model and variable-inclusion summaries と操作
- `bmastats models` ダイアログ: デフォルト設定のまま実行

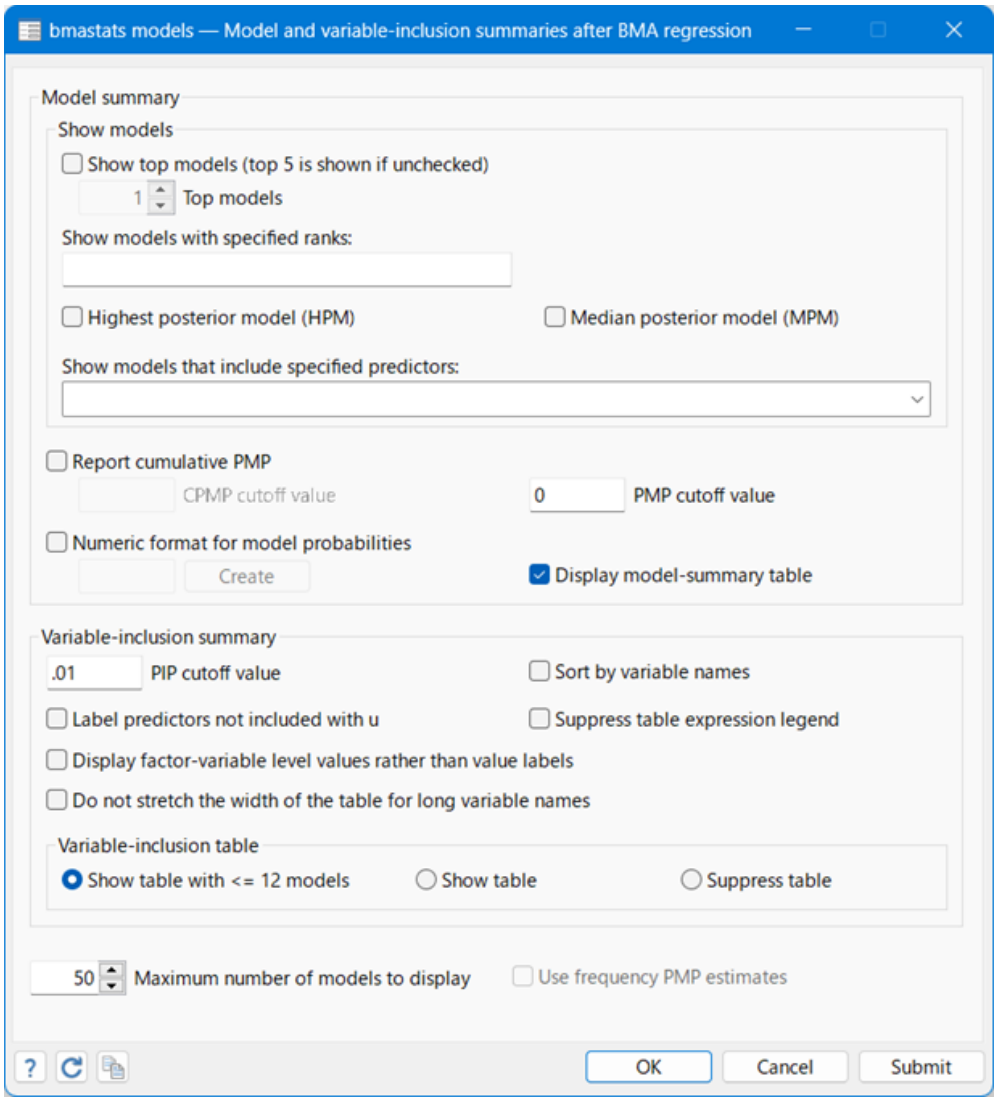

図 2 `bmastats models` ダイアログ

```
. bmastats models

Computing model probabilities ...

Model summary         Number of models:
                               Visited = 1,024
                              Reported =     5
```

	Analytical PMP	Model size
Rank		
1	.6292	2
2	.1444	3
3	.0258	3
4	.0246	3
5	.01996	3

Variable-inclusion summary

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
x2	x	x	x	x	x
x10	x	x	x	x	x
x3		x			
x9			x		
x1				x	
x4					x

```
Legend:
  x - estimated
```

予想通り、x2 と x10 からなるモデルがランク 1 に位置付けられており、その PMP は 0.63 とレポートされています。ランク 2 のモデルはそれに x3 を加えたモデルであるわけですが、PMP は 0.14 とはるかに小さな値となっています。

上での操作において、bmaregress はデフォルトの事前分布 (priors) を使用しています（bmaregress 出力中のヘッダ部を参照)。これらの事前分布はユーザの便宜を図るために提供されているものであって、本来はそれぞれのアプリケーションごとに評価されるべきものです。特に priors を変えたときにどのような影響が出るかを評価する感度解析 (sensitivity analysis) は大切です。この点については [BMA] **bmaregress** (*mwp-470*) の Example 11 を参照ください。

BMA においては、回帰係数に対する prior の分散はいわゆる g パラメータに比例したものとなります。デフォルトの場合、g は固定値 $\max(n, p^2)$ を取るわけですが、ここでの例では $g = n = 200$ となります。これを例えば 1,000 に変更すれば分散に対する制約を緩和することができ、より OLS の推定値に近い結果を得ることができます。

BMA のもう 1 つのメリットはモデル用の prior を介してモデル不確定性を制御し得るという点です。例えば仮に予測変数 `x1, x3-x9` が `y` とは関係しないという事前の知識を持っていたとするなら、それを BMA モデルの中に組み入れることが可能です。次に示す例においては `mprior()` オプションを使い二項のモデル prior を設定しているわけですが、その際、`x2` と `x10` に対しては含有確率 (inclusion probability) `0.5` を、`x1, x3-x9` に対しては含有確率 `0.1` を指定しています。

- `bmaregress` ダイアログ:
  - Model タブ: Dependent variable: `y`
    - In-out predictors: `x1-x10`
    - Prior for model space: Binomial with different inclusion probabilities
      - In-out terms: In-out terms: `x2 x10` Probability: `0.5` ⇒ Add
      - In-out terms: `x1 x3-x9` Probability: `0.1` ⇒ Add

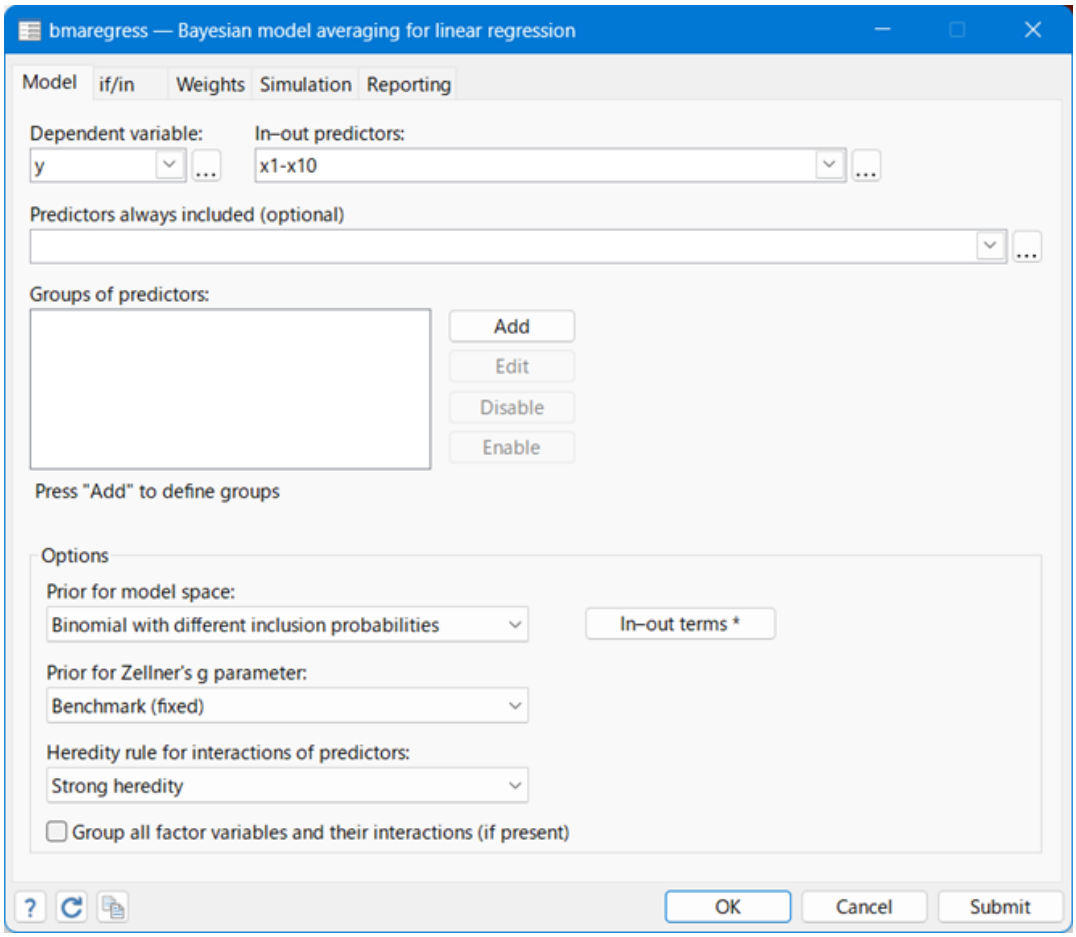

図 3 `bmaregress` ダイアログ - Model タブ

```
. bmaregress y x1-x10, mprior(binomial x2 x10 0.5  x1 x3-x9 0.1)

Enumerating models ...
Computing model probabilities ...

Bayesian model averaging                        No. of obs         =    200
Linear regression                               No. of predictors  =     10
Model enumeration                                           Groups =     10
                                                            Always =      0
Priors:                                         No. of models      =  1,024
  Models: Binomial, IP varies                       For CPMP >= .9 =      2
   Cons.: Noninformative                        Mean model size    =  2.129
   Coef.: Zellner's g
       g: Benchmark, g = 200                    Shrinkage, g/(1+g) = 0.9950
  sigma2: Noninformative                        Mean sigma2        =  1.276

------------------------------------------------------------------------------
           y |       Mean   Std. dev.                      Group        PIP
-------------+----------------------------------------------------------------
          x2 |   1.200944   .0730381                          2          1
         x10 |   5.080663   .0899736                         10          1
          x3 |  -.0106068   .0452704                          3    .064039
          x9 |   .0009677   .0126993                          9    .012195
          x1 |   .0008208   .0115323                          1     .01149
-------------+----------------------------------------------------------------
Always       |
       _cons |   .5884159   .0803504                          0          1
------------------------------------------------------------------------------
Note: Coefficient posterior means and std. dev. estimated from 1,024 models.
Note: Default prior is used for parameter g.
Note: 5 predictors with PIP less than .01 not shown.
```

このモデル prior の効果は予測変数 x1 と x3–x9 の PIP がいずれも 8% よりも小さな値となっていることです。また切片項と x2 の係数推定値に若干の改善が見られます。

科学や過去の知見に基づく事前の前提条件をモデル中に入れられるという特質は標準的なベイズ分析に由来するものです。そのような事前分布が存在する場合に、BMA フレームワークは古典的な回帰手法よりも信頼度の高い推論を誘導できる潜在能力を秘めています。 ◁

▷ Example 2: BMA **予測性能**

評価版では割愛しています。

■