

第十一章 贝叶斯模型平均

Wang Shujia

Department of Statistics, School of Economics
Shenzhen University



Outline

- 1 模型的不确定性
- 2 贝叶斯模型平均法基本原理
- 3 BMA 法在肥胖指数预测中的应用
- 4 BMA 软件与运用
- 5 讨论与结论
- 6 参考文献

美国 47 个州的犯罪率问题

研究目的：犯罪率的影响因素。

- 数据：Ehrlich (1973)；数据修正：Vandaele (1978)
- 犯罪理论：
 - ▶ 60 年前：罪犯的心理、家庭和社会环境造成
 - ▶ Ehrlich：犯罪成本与收益的理性选择

共 15 个自变量

- 监禁的概率、服刑平均时间、警察费用支出
- 社区的富裕程度、收入不平等程度
- 受教育程度、就业难易程度
- 非白人比例、年轻男性比例
- 控制变量 (Vandaele, 1978): 是否南方州、性别比例、州总人口

自变量

#	Variable
1	percent of males 14-24
2	indicator variable for southern state
3	mean years of schooling
4	police expenditure in 1960
5	police expenditure in 1959
6	labor force participation rate
7	number of males per 1000 females
8	state population
9	number of nonwhites per 1000 people
10	unemployment rate of urban males 14-24
11	unemployment rate of urban males 35-39
12	GDP
13	income inequality
14	probability of imprisonment
15	average time served in state prisons

模型结果比较

- 所有变量均取对数
- 比较 6 个模型：完整、逐步回归、 C_p 、修正的 R^2 、Ehrlich1、Ehrlich2
- Ehrlich 是依据其理论预先设定的，没有按统计方法选取
- 重点比较变量 14（犯罪被监禁的概率）和 15（平均刑期）

#	Method	Variables	R^2 (%)	# vars.	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$	P_{15}
1	Full model	All	87	15	-.30	-.27	.133
2	Stepwise regression	1,3,4,9,11,13,14	83	7	-.19	—	—
3	Mallows' C_p	1,3,4,9,11,12,13,14,15	85	9	-.30	-.30	.050
4	Adjusted R^2	1,3,4,7,8,9,11,12,13,14,15	86	11	-.30	-.25	.129
5	Ehrlich model 1	9,12,13,14,15	66	5	-.45	-.55	.009
6	Ehrlich model 2	1,6,9,10,12,13,14,15	70	8	-.43	-.53	.011

P_{15} 是 $\beta_{15} = 0$ 的 t-检验的 p-值。

模型结果的描述

① β_{14}

- ▶ 完整模型与逐步回归模型结果相差 $1/3$ ，差值相当于 1.7 个标准误。
- ▶ Ehrlich 模型与逐步回归模型结果相差超过一倍
- ▶ 不同模型中系数估计值的差异大到需要改变政策：模型不确定性很大

模型结果的描述

① β_{14}

- ▶ 完整模型与逐步回归模型结果相差 1/3，差值相当于 1.7 个标准误。
- ▶ Ehrlich 模型与逐步回归模型结果相差超过一倍
- ▶ 不同模型中系数估计值的差异大到需要改变政策：模型不确定性很大

② β_{15}

- ▶ 完整模型：不显著 ($p=0.133$)
- ▶ 逐步回归：被剔除
- ▶ 修正的 R^2 ：不显著
- ▶ Ehrlich 两个模型：高度显著
 - ★ 与其他模型不一致，需要更强的理论支持
- ▶ 不同模型下，影响因素的显著性出现矛盾

模型结果的描述

① β_{14}

- ▶ 完整模型与逐步回归模型结果相差 1/3，差值相当于 1.7 个标准误。
- ▶ Ehrlich 模型与逐步回归模型结果相差超过一倍
- ▶ 不同模型中系数估计值的差异大到需要改变政策：模型不确定性很大

② β_{15}

- ▶ 完整模型：不显著 ($p=0.133$)
 - ▶ 逐步回归：被剔除
 - ▶ 修正的 R^2 ：不显著
 - ▶ Ehrlich 两个模型：高度显著
 - ★ 与其他模型不一致，需要更强的理论支持
 - ▶ 不同模型下，影响因素的显著性出现矛盾
- 结论：根据不同理论、方法和标准选择的“最优”模型，存在不确定性，可能得出差异很大甚至相互矛盾的结论
 - 办法：对所有可能的模型（或若干重要的模型）加以平均，权重为各模型的后验概率
 - 优点：克服模型不确定性，参数估计或预测具有更大的稳健性

Outline

- 1 模型的不确定性
- 2 贝叶斯模型平均法基本原理**
- 3 BMA 法在肥胖指数预测中的应用
- 4 BMA 软件与运用
- 5 讨论与结论
- 6 参考文献

基本记号

- Δ : 感兴趣的量 (回归系数、未来观察值等, 如 β_{15})
- D : 数据
- $\mathcal{M} = \{M_k, k = 1, \dots, K\}$ 所有候选模型, 共有 2^K 个
- θ_k : 模型 M_k 的参数向量
- $p(\theta_k|M_k)$: 给定模型 M_k 下, θ_k 的先验分布密度
- $f(D|\theta_k, M_k)$: 给定模型 M_k 和参数 θ_k 下的概率模型 (似然函数)
- $p(M_k)$: M_k 模型的先验概率

- 2 贝叶斯模型平均法基本原理
 - 一般参数的 BMA 估计
 - 解释变量系数的 BMA 估计
 - BMA 的实施
 - 先验分布
 - 模型的预测表现

贝叶斯模型平均法

给定数据 D , Δ 的后验分布为:

$$p(\Delta|D) = \sum_{k=1}^K p(\Delta|D, M_k)p(M_k|D) \quad (1)$$

其中 $p(M_k|D)$ 为 M_k 的后验模型概率 (PMP, posterior model probability), 计算公式为

$$p(M_k|D) = \frac{p(D|M_k)p(M_k)}{\sum_{i=1}^K p(D|M_i)p(M_i)} \quad (2)$$

其中

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (3)$$

- Δ 的后验分布等于其在每个模型下的后验分布的加权平均, 权重为后验模型概率 (PMP)

后验均值与后验方差

- 记 $\hat{\Delta}_k = E(\Delta|D, M_k)$ (模型 M_k 下 Δ 的后验均值)

后验均值与后验方差

- 记 $\hat{\Delta}_k = E(\Delta|D, M_k)$ (模型 M_k 下 Δ 的后验均值)
- $SE_k^2 = \text{Var}[\Delta|D, M_k]$

后验均值与后验方差

- 记 $\hat{\Delta}_k = E(\Delta|D, M_k)$ (模型 M_k 下 Δ 的后验均值)
- $SE_k^2 = \text{Var}[\Delta|D, M_k]$
- 则 Δ 的后验均值和后验方差为

$$E(\Delta|D) = \sum_{k=1}^K \hat{\Delta}_k p(M_k|D) \quad (4)$$

$$\text{Var}(\Delta|D) = \sum_{k=1}^K (SE_k^2 + \hat{\Delta}_k^2) p(M_k|D) - E[\Delta|D]^2 \quad (5)$$

- 2 贝叶斯模型平均法基本原理
 - 一般参数的 BMA 估计
 - 解释变量系数的 BMA 估计
 - BMA 的实施
 - 先验分布
 - 模型的预测表现

- 在所有候选模型中，有部分模型的 $\beta_1 = 0$ ，因此后验概率 $P(\beta_1 = 0|D) \neq 0$

变量入选概率

- 在所有候选模型中，有部分模型的 $\beta_1 = 0$ ，因此后验概率 $P(\beta_1 = 0|D) \neq 0$
- 变量 β_1 入选模型的后验概率 (PIP:posterior inclusion probabilities or PEP:posterior effect probabilities)

$$P(\beta_1 \neq 0|D) = \sum_{\mathcal{A}_1} p(M_k|D) \quad (6)$$

其中 $\mathcal{A}_1 = \{M_k : k = 1, \dots, K; \beta_1 \neq 0\}$ 。

变量入选概率

- 在所有候选模型中，有部分模型的 $\beta_1 = 0$ ，因此后验概率 $P(\beta_1 = 0|D) \neq 0$
- 变量 β_1 入选模型的后验概率 (PIP:posterior inclusion probabilities or PEP:posterior effect probabilities)

$$P(\beta_1 \neq 0|D) = \sum_{\mathcal{A}_1} p(M_k|D) \quad (6)$$

其中 $\mathcal{A}_1 = \{M_k : k = 1, \dots, K; \beta_1 \neq 0\}$ 。

- 即 PIP 等于所有包含 β_1 的后验模型概率 (PMP) 之和

系数的 BMA 估计

- β_1 的后验分布为

$$p(\beta_1|D, \beta_1 \neq 0) = \sum_{\mathcal{A}_1} p(\beta_1|D, M_k)p(M_k|D) \quad (7)$$

- β_1 的后验均值和后验方差为

$$E(\beta_1|D, \beta_1 \neq 0) = \sum_{\mathcal{A}_1} \hat{\beta}_1(k)p(M_k|D) \quad (8)$$

$$\text{Var}(\beta_1|D, \beta_1 \neq 0) = \sum_{\mathcal{A}_1} [\text{SE}(\beta_1)_k]^2 + [\hat{\beta}_1(k)]^2)p(M_k|D) - E[\Delta|D]^2 \quad (9)$$

其中 $\hat{\beta}_1(k)$ 和 $\text{SE}(\beta_1)_k$ 分别表示模型 M_k 下 β_1 的估计量和标准误差。

- 2 贝叶斯模型平均法基本原理
 - 一般参数的 BMA 估计
 - 解释变量系数的 BMA 估计
 - **BMA 的实施**
 - 先验分布
 - 模型的预测表现

- ① 模型平均： \mathcal{M} 数量可能巨大（ $q = 15$ ，则有 $2^{15} = 32768$ 个模型）

$$p(\Delta|D) = \sum_{k=1}^K p(\Delta|D, M_k)p(M_k|D) \quad (10)$$

BMA 实施需解决的关键问题

- ① 模型平均： \mathcal{M} 数量可能巨大（ $q = 15$ ，则有 $2^{15} = 32768$ 个模型）

$$p(\Delta|D) = \sum_{k=1}^K p(\Delta|D, M_k)p(M_k|D) \quad (10)$$

- ② 积分计算：维度大，即使用 MCMC 也有难度

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (11)$$

BMA 实施需解决的关键问题

- ① 模型平均： \mathcal{M} 数量可能巨大（ $q = 15$ ，则有 $2^{15} = 32768$ 个模型）

$$p(\Delta|D) = \sum_{k=1}^K p(\Delta|D, M_k)p(M_k|D) \quad (10)$$

- ② 积分计算：维度大，即使用 MCMC 也有难度

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (11)$$

- ③ 模型的先验分布难以合理给定

BMA 实施需解决的关键问题

- ① 模型平均： \mathcal{M} 数量可能巨大（ $q = 15$ ，则有 $2^{15} = 32768$ 个模型）

$$p(\Delta|D) = \sum_{k=1}^K p(\Delta|D, M_k)p(M_k|D) \quad (10)$$

- ② 积分计算：维度大，即使用 MCMC 也有难度

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (11)$$

- ③ 模型的先验分布难以合理给定
- ④ 选取哪些模型进行平均需抉择

模型平均的方法：奥卡姆剃刀和 MC^3

方法一：奥卡姆窗（Occam's window）(Madigan and Raftery, 1994)

模型平均的方法：奥卡姆剃刀和 MC^3

方法一：奥卡姆窗（Occam's window）(Madigan and Raftery, 1994)

根据如下两个原则，剔除部分模型：

- ① 剔除预测能力低的模型（如 BIC 相差 6）
 - ② 奥卡姆剃刀（Occam's Razor）：如果数据支持简单模型，则应剔除复杂模型
- 奥卡姆窗称为对称的（symmetric）：如果满足 (1)
 - 奥卡姆窗称为严格的（strict）：如果满足 (1) 和 (2)
 - 美国犯罪率例子：对称奥卡姆窗有 51 个模型，严格奥卡姆窗仅 14 个模型

模型平均的方法：奥卡姆剃刀和 MC^3

方法一：奥卡姆窗（Occam's window）(Madigan and Raftery, 1994)

根据如下两个原则，剔除部分模型：

- ① 剔除预测能力低的模型（如 BIC 相差 6）
 - ② 奥卡姆剃刀（Occam's Razor）：如果数据支持简单模型，则应剔除复杂模型
- 奥卡姆窗称为对称的（symmetric）：如果满足 (1)
 - 奥卡姆窗称为严格的（strict）：如果满足 (1) 和 (2)
 - 美国犯罪率例子：对称奥卡姆窗有 51 个模型，严格奥卡姆窗仅 14 个模型

方法二：Markov chain Monte Carlo model composition (MC^3)

参看 Koop, G. (2003): Bayesian Econometrics, Wiley-Interscience(ch11).

- BMA 的另一难点是需要计算如下积分:

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (12)$$

- 方法:

- ▶ 对线性回归模型, 积分可以求出显式解
- ▶ Laplace 方法 (Tierney and Kadane,1986): 可以很好近似 $p(D|M_k)$, 而且某些条件下可以给出 BIC 近似值
- ▶ MLE 近似法 (Taplin (1993)): 用 $p(\Delta|D, \hat{\theta}, M_k)$ 代替 $p(\Delta|D, M_k)$

- 2 贝叶斯模型平均法基本原理
 - 一般参数的 BMA 估计
 - 解释变量系数的 BMA 估计
 - BMA 的实施
 - 先验分布
 - 模型的预测表现

模型的先验概率

- ① 均匀先验: 指定每个模型相同的先验概率, $P(M_j) = 2^{-q}$
- ② Binomial Priors: 设模型规模为 N (入选解释变量个数), 每个变量入选的先验概率为 π , 则 $N \sim Bin(q, \pi)$, 其中 q 为纳入研究的全部自变量个数。
 - ▶ 模型 M_i (有 q_i 个自变量) 的先验概率为

$$P(M_i) = \pi^{q_i} (1 - \pi)^{q - q_i} \quad (13)$$

- ▶ 平均模型规模为 $E(N) = q\pi$
 - ▶ 若假设 $\pi = 0.5$ (均匀分布), 则 $E(N) = q/2$, 此时等价于均匀先验。
 - ▶ Package BMS: `mprior="fixed"`
- ③ Binomial-Beta Priors: Ley and Steel (2009)

$$N \sim Bin(q, \pi), \pi \sim Beta(a, b) \quad (14)$$

- ▶ Package BMS: `mprior="random"` (Default)
- ④ Custom Prior Inclusion Probabilities: 给每个变量指定一个入选概率。

模型系数的先验分布

假设线性回归模型

$$y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2 I_T) \quad (15)$$

其中

- $y = (y_1, y_2, \dots, y_T)'$
- X 为 $T \times q$ 矩阵 (包含所有可能变量)
- $\beta = (\beta_1, \beta_2, \dots, \beta_q)'$, 设部分分量为零, 得到所有候选模型 $M_j, j = 1, \dots, K = 2^q$
- 模型 M_j 对应的参数记为 β^j

模型系数的先验分布

- Zellner's g-Priors(Zellner (1986))

$$\beta^j | \sigma^2, M_j, g \sim N(0, \sigma^2 g (X_j' X_j)^{-1}), p(\sigma) \propto \sigma^{-1}$$

- ▶ Unit Information Prior (g-UIP): $g = N$, 样本容量, 其贝叶斯因子与 BIC 相似 (Kass and Wasserman (1995))
 - ▶ Risk Inflation Criterion (g-RIC): $g = q^2$ (Foster and George (1994))
 - ▶ Benchmark Prior: $g = \max(N; q^2)$ (Fernández et al. (2001))
- Empirical Bayes Priors
 - Laplace Priors

- 2 贝叶斯模型平均法基本原理
 - 一般参数的 BMA 估计
 - 解释变量系数的 BMA 估计
 - BMA 的实施
 - 先验分布
 - 模型的预测表现

模型的预测能力：对数预测得分

- 数据随机地分为两半：训练集 (training) 和测试集 (testing)
- 样本内准确率和样本外准确率
- 对数预测得分 (Log Predictive Score, LPS) (Good, (1952)):

- ▶ 单个模型:

$$\sum_{d \in D^{\text{test}}} \{ -\log p(d|M, D^{\text{train}}) \} \quad (16)$$

- ▶ BMA:

$$\sum_{d \in D^{\text{test}}} \{ -\log \{ \sum_{M \in \mathcal{A}} p(d|M, D^{\text{train}}) p(M|D^{\text{train}}) \} \} \quad (17)$$

- ▶ LPS 越小表示预测表现越好

模型的预测能力：交叉验证及信息准则

① 交叉验证 (Cross Validation)

- ① K -折交叉验证 (k-fold cross validation): 将数据分成 K 份 (尽可能每份数据个数相同), 用其中 $(K - 1)$ 份训练模型, 剩下一份数据用于验证模型。重复 K 轮, 每轮用不同数据进行模型验证, 最后对每一轮的验证结果进行平均。
- ② 留一交叉验证 (Leave-One-Out cross validation): K 与样本个数相同, 即每次仅留一个数据用于验证模型。

② 信息准则: AIC、BIC、DIC 等

- ▶ 数据很大时, 交叉验证计算量太大, 所以用信息准则近似
- ▶ 对 BMA 而言, 贝叶斯更喜欢用对数预测得分 (LPS), 因为 LPS 来源于每个模型的后验分布

Outline

- 1 模型的不确定性
- 2 贝叶斯模型平均法基本原理
- 3 BMA 法在肥胖指数预测中的应用**
- 4 BMA 软件与运用
- 5 讨论与结论
- 6 参考文献

研究目的

- 研究目的是利用 13 个身体指标预测肥胖指数 (percent body fat)
- 比较 BMA 与基于标准方法选择的单一模型
 - ▶ 模型选择
 - ▶ 预测表现

TABLE 6
Body fat example: summary statistics for full data set¹

Predictor number	Predictor	mean	s.d.	min	max
X_1	Age (years)	45	13	21	81
X_2	Weight (pounds)	179	29	118	363
X_3	Height (inches)	70	3	64	78
X_4	Neck circumference (cm)	38	2	31	51
X_5	Chest circumference (cm)	101	8	79	136
X_6	Abdomen circumference (cm)	93	11	69	148
X_7	Hip circumference (cm)	100	7	85	148
X_8	Thigh circumference (cm)	59	5	47	87
X_9	Knee circumference (cm)	39	2	33	49
X_{10}	Ankle circumference (cm)	23	2	19	34
X_{11}	Extended biceps circumference	32	3	25	45
X_{12}	Forearm circumference (cm)	29	2	21	35
X_{13}	Wrist circumference (cm)	18	1	16	21

¹Abdomen circumference was measured at the umbilicus and level with the iliac crest. Wrist circumference (cm) was measured distal to the styloid processes.

对比模型

- 模型 1: 完整模型
- 模型 2: 逐步回归;
- 模型 3: 最小 Mallows's C_p ;
- 模型 4: 最大 adjusted R^2
- BMA: 无信息先验 (模型先验概率相同)
 - ▶ 有 13 个变量, 共有 $2^{13} = 8192$ 个候选模型

TABLE 7
*Body fat example: least squares regression results
 from the full model¹*

Predictor		Coef	Std error	t-statistic	p-value
Intercept		-17.80	20.60	-0.86	0.39
X_1	age	0.06	0.03	1.89	0.06
X_2	weight	-0.09	0.06	-1.50	0.14
X_3	height	-0.04	0.17	-0.23	0.82
X_4	neck	-0.43	0.22	-1.96	0.05
X_5	chest	-0.02	0.10	-0.19	0.85
X_6	abdomen	0.89	0.08	10.62	<0.01
X_7	hip	-0.20	0.14	-1.44	0.15
X_8	thigh	0.24	0.14	1.74	0.08
X_9	knee	-0.02	0.23	-0.09	0.93
X_{10}	ankle	0.17	0.21	0.81	0.42
X_{11}	biceps	0.16	0.16	0.98	0.33
X_{12}	forearm	0.43	0.18	2.32	0.02
X_{13}	wrist	-1.47	0.50	-2.97	<0.01

¹Residual standard error = 4, $R^2 = 0.75$, $N = 251$, F -statistic = 53.62 on 13 and 237 df, p -value <0.0001.

显著: 4(0.05), 6(<0.01), 12(0.02), 13(<0.01)

比较：BMA 与逐步回归模型

TABLE 8

Body fat example: comparison of BMA results to model selected using standard model selection methods¹

Predictor		Bayesian model averaging			Stepwise model
		Mean βD	SD βD	$P(\beta \neq 0 D)$	p-value
X_6	abdomen	1.2687	0.08	100	<0.01
X_2	weight	-0.4642	0.15	97	0.03
X_{13}	wrist	-0.0924	0.08	62	<0.01
X_{12}	forearm	0.0390	0.06	35	0.01
X_4	neck	-0.0231	0.06	19	0.05
X_{11}	biceps	0.0179	0.05	17	
X_8	thigh	0.0176	0.05	15	0.02
X_7	hip	-0.0196	0.07	13	0.12
X_5	chest	0.0004	0.02	6	
X_1	age	0.0029	0.02	5	0.05
X_9	knee	0.0020	0.02	5	
X_3	height	-0.0015	0.01	4	
X_{10}	ankle	0.0011	0.01	4	

¹Stepwise, minimum Mallows's C_p , and maximum adjusted R^2 all selected the same model. The predictors are sorted by $P(\beta_i \neq 0|D)$ which is expressed as a percentage. The results given here are based on standardized data (columns have means equal to 0 and variances equal to 1).

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:
 - ▶ PIP 排前三的变量 (Weight, Abdomen, and Wrist): 逐步回归都显著

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:
 - ▶ PIP 排前三的变量 (Weight, Abdomen, and Wrist): 逐步回归都显著
- p-值高估了显著性:

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:
 - ▶ PIP 排前三的变量 (Weight, Abdomen, and Wrist): 逐步回归都显著
- p-值高估了显著性:
 - ▶ forearm, neck and thigh: p-值显著 (<0.05), 但 PIP 表明显著性不明显 (PIP 在 15-35 之间)

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:
 - ▶ PIP 排前三的变量 (Weight, Abdomen, and Wrist): 逐步回归都显著
- p-值高估了显著性:
 - ▶ forearm, neck and thigh: p-值显著 (<0.05), 但 PIP 表明显著性不明显 (PIP 在 15-35 之间)
 - ▶ age: p-值显著 (0.05), 但 $PIP(=5)$ 不显著, 结论相反

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:
 - ▶ PIP 排前三的变量 (Weight, Abdomen, and Wrist): 逐步回归都显著
- p -值高估了显著性:
 - ▶ forearm, neck and thigh: p -值显著 (<0.05), 但 PIP 表明显著性不明显 (PIP 在 15-35 之间)
 - ▶ age: p -值显著 (0.05), 但 $PIP(=5)$ 不显著, 结论相反
- 与完整模型结论有冲突:

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:
 - ▶ PIP 排前三的变量 (Weight, Abdomen, and Wrist): 逐步回归都显著
- p-值高估了显著性:
 - ▶ forearm, neck and thigh: p-值显著 (<0.05), 但 PIP 表明显著性不明显 (PIP 在 15-35 之间)
 - ▶ age: p-值显著 (0.05), 但 PIP(=5) 不显著, 结论相反
- 与完整模型结论有冲突:
 - ▶ Weight 在 BMA、逐步回归模型均显著, 但在完整模型中不显著 (p-值 =0.14)

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:
 - ▶ PIP 排前三的变量 (Weight, Abdomen, and Wrist): 逐步回归都显著
- p-值高估了显著性:
 - ▶ forearm, neck and thigh: p-值显著 (<0.05), 但 PIP 表明显著性不明显 (PIP 在 15-35 之间)
 - ▶ age: p-值显著 (0.05), 但 PIP(=5) 不显著, 结论相反
- 与完整模型结论有冲突:
 - ▶ Weight 在 BMA、逐步回归模型均显著, 但在完整模型中不显著 (p-值 =0.14)
- 不同模型之间系数估计值差异较大

结果：部分变量得到不同结论

- $P(\beta_i \neq 0|D)$: PIP or PEP, 所有包含第 i 个变量的后验模型概率 (PMP) 之和
- 主要变量比较一致:
 - ▶ PIP 排前三的变量 (Weight, Abdomen, and Wrist): 逐步回归都显著
- p-值高估了显著性:
 - ▶ forearm, neck and thigh: p-值显著 (<0.05), 但 PIP 表明显著性不明显 (PIP 在 15-35 之间)
 - ▶ age: p-值显著 (0.05), 但 PIP(=5) 不显著, 结论相反
- 与完整模型结论有冲突:
 - ▶ Weight 在 BMA、逐步回归模型均显著, 但在完整模型中不显著 (p-值 =0.14)
- 不同模型之间系数估计值差异较大
- 结论: 存在模型不确定性

PMP: 反映模型不确定性

TABLE 9
Body fat example: Ten models with highest posterior model probability (PMP)

X_2	X_4	X_6	X_8	X_{11}	X_{12}	X_{13}	PMP
•		•				•	0.14
•		•			•	•	0.14
•		•					0.12
•		•		•		•	0.05
•		•	•				0.03
•	•	•					0.03
•		•			•		0.02
•		•		•			0.02
•	•	•			•	•	0.02
•	•	•			•		0.02

- 存在较大的模型不确定性：PMP 最大的模型仅占总后验概率的 14%，排前 10 的模型 PMP 之和仅 57%

变量 Wrist 系数（效应）的后验分布

- β_{13} 的后验分布, $P(\beta_{13} = 0|D) = 0.38$

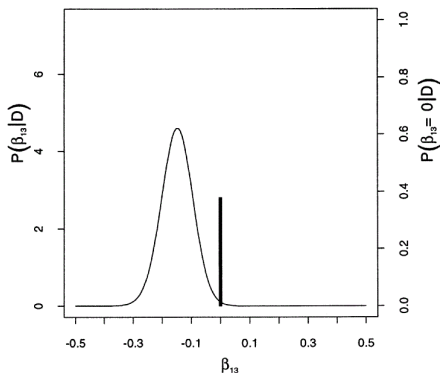


FIG. 4. *Body fat example: BMA posterior distribution for β_{13} , the coefficient for wrist circumference. The spike corresponds to $P(\beta_{13} = 0|D) = 0.38$. The vertical axis on the left corresponds to the posterior distribution for β_{13} and the vertical axis on the right corresponds to the posterior distribution for β_{13} equal to 0. The density is scaled so that the maximum of the density is equal to $P(\beta_{13} \neq 0|D)$ on the right axis.*

预测表现 (Predictive Performance)

- 数据分为两部分，前 142 个数据用于建模，后 109 个数据用于比较模型预测的准确度

TABLE 10
*Body fat example: performance comparison*¹

Method	Model	Predictive coverage %
BMA	Model averaging	90.8
Stepwise and C_p	$X_1 X_2 X_6 X_{10} X_{12} X_{13}$	84.4
Adjusted R^2	$X_1 X_2 X_4 X_6 X_7 X_8 X_{10} X_{12} X_{13}$	83.5

¹Predictive coverage % is the percentage of observations in the performance set that fall in the 90% prediction interval. For BMA, the top 2500 models, accounting for 99.99% of the posterior model probability, were used to estimate predictive coverage.

预测表现 (Predictive Performance)

- 数据分为两部分，前 142 个数据用于建模，后 109 个数据用于比较模型预测的准确度
- 预测区间：基于单个模型（三个不同标准选出的模型不一样）和 BMA 法的后验预测分布

TABLE 10
*Body fat example: performance comparison*¹

Method	Model	Predictive coverage %
BMA	Model averaging	90.8
Stepwise and C_p	$X_1 X_2 X_6 X_{10} X_{12} X_{13}$	84.4
Adjusted R^2	$X_1 X_2 X_4 X_6 X_7 X_8 X_{10} X_{12} X_{13}$	83.5

¹Predictive coverage % is the percentage of observations in the performance set that fall in the 90% prediction interval. For BMA, the top 2500 models, accounting for 99.99% of the posterior model probability, were used to estimate predictive coverage.

预测表现 (Predictive Performance)

- 数据分为两部分，前 142 个数据用于建模，后 109 个数据用于比较模型预测的准确度
- 预测区间：基于单个模型（三个不同标准选出的模型不一样）和 BMA 法的后验预测分布
- 预测覆盖率：实际数据落在 90% 预测区间的比例

TABLE 10
*Body fat example: performance comparison*¹

Method	Model	Predictive coverage %
BMA	Model averaging	90.8
Stepwise and C_p	$X_1 X_2 X_6 X_{10} X_{12} X_{13}$	84.4
Adjusted R^2	$X_1 X_2 X_4 X_6 X_7 X_8 X_{10} X_{12} X_{13}$	83.5

¹Predictive coverage % is the percentage of observations in the performance set that fall in the 90% prediction interval. For BMA, the top 2500 models, accounting for 99.99% of the posterior model probability, were used to estimate predictive coverage.

预测表现 (Predictive Performance)

- 数据分为两部分，前 142 个数据用于建模，后 109 个数据用于比较模型预测的准确度
- 预测区间：基于单个模型（三个不同标准选出的模型不一样）和 BMA 法的后验预测分布
- 预测覆盖率：实际数据落在 90% 预测区间的比例
- 结论：BMA 预测覆盖率 90.8%，明显由于单一模型 (<85%)

TABLE 10
*Body fat example: performance comparison*¹

Method	Model	Predictive coverage %
BMA	Model averaging	90.8
Stepwise and C_p	$X_1 X_2 X_6 X_{10} X_{12} X_{13}$	84.4
Adjusted R^2	$X_1 X_2 X_4 X_6 X_7 X_8 X_{10} X_{12} X_{13}$	83.5

¹Predictive coverage % is the percentage of observations in the performance set that fall in the 90% prediction interval. For BMA, the top 2500 models, accounting for 99.99% of the posterior model probability, were used to estimate predictive coverage.

预测表现 (Predictive Performance)

- 数据分为两部分，前 142 个数据用于建模，后 109 个数据用于比较模型预测的准确度
- 预测区间：基于单个模型（三个不同标准选出的模型不一样）和 BMA 法的后验预测分布
- 预测覆盖率：实际数据落在 90% 预测区间的比例
- 结论：BMA 预测覆盖率 90.8%，明显由于单一模型 (<85%)
- 对数据进行不同划分，选取的模型会不同，但基本上结论一致

TABLE 10
*Body fat example: performance comparison*¹

Method	Model	Predictive coverage %
BMA	Model averaging	90.8
Stepwise and C_p	$X_1 X_2 X_6 X_{10} X_{12} X_{13}$	84.4
Adjusted R^2	$X_1 X_2 X_4 X_6 X_7 X_8 X_{10} X_{12} X_{13}$	83.5

¹Predictive coverage % is the percentage of observations in the performance set that fall in the 90% prediction interval. For BMA, the top 2500 models, accounting for 99.99% of the posterior model probability, were used to estimate predictive coverage.

Outline

- 1 模型的不确定性
- 2 贝叶斯模型平均法基本原理
- 3 BMA 法在肥胖指数预测中的应用
- 4 BMA 软件与运用**
- 5 讨论与结论
- 6 参考文献

- The BMS Package
 - ▶ The BMS (Bayesian Model Selection) package employs standard Bayesian normal-conjugate linear model as the base model and Zellner's g prior as the choice of prior structures for the regression coefficients (Feldkircher & Zeugner 2009).
- The BAS Package
 - ▶ The BAS package (Bayesian Adaptive Sampling), performs BMA in linear models using stochastic or deterministic sampling without replacement from posterior distributions (Clyde 2010).
- The BMA Package
 - ▶ The BMA package (Bayesian model averaging), performs BMA analysis assuming a uniform distribution on model priors and using a simple BIC approximation to construct the prior probabilities on the regressions coefficients (Raftery, Hoeting, Volinsky, Painter & Yeung 2010).
- <https://cran.r-project.org/web/views/Bayesian.html>

数据

数据网址: <https://ww2.amstat.org/publications/jse/datasets/fat.dat.txt>

变量说明: <https://ww2.amstat.org/publications/jse/datasets/fat.txt>

解释变量: columns 5-7 and 10-19

因变量: 要放在数据的首列

剔除第 42 行, 因为身高仅 29.5 英寸 (75cm)。

```
fat1=read.table("fat.dat.txt")
fat = fat1[,c(2, 5:7, 10:19)]
names(fat)=c("brozek_fat","age","weight","height",
             "neck","chest","abdomen","hip","thigh",
             "knee","ankle","biceps","forearm","wrist")
fat= fat[-42,]
```

BMS 计算代码

```
library(BMS)
mf= bms(fat,user.int=F) # 默认选项, 非交互输出
coef(mf,std.coefs=T,include.constant=F) # 系数标准化, 不包含常数项
topmodels.bma(mf)[, 1:5] # 排在前 5 位模型及 PMP
density(mf,reg="wrist",std.coefs=T) # 边缘后验密度
summary(mf) # 显示模型综合信息
image(mf) # 图示模型入选情况
plotModelsize(mf) # 模型规模的先后验对比图
plotConv(mf) # 比较 PMP:MCMC 模型与实际模型
```


BMA 预测代码

```
# 用前 250 个数据预测最后一个数据
mf.p = bms(fat[1:250,], user.int = F)
# 预测分布密度
dens.p = pred.density(mf.p,newdata=fat[251,])
plot(dens.p)
# 预测区间
quantile(dens.p,c(0.05,0.95))
# 预测表现 LPS
lps.bma(dens.p,realized.y=fat[251,1],newdata=fat[251,-1])
# 用前 200 个数据预测后 51 个?
```

BMA 系数估计及入选概率

```
> round(coef(mf, std.coefs = T, include.constant = F), 3)
      PIP Post Mean Post SD Cond.Pos.Sig. Idx
abdomen 1.000    1.265  0.083      1.000   6
weight  0.965   -0.471  0.143      0.000   2
wrist   0.527   -0.079  0.085      0.000  13
forearm 0.306    0.034  0.057      1.000  12
neck    0.135   -0.017  0.050      0.000   4
biceps  0.124    0.013  0.041      1.000  11
thigh   0.094    0.011  0.041      0.999   8
hip     0.071   -0.013  0.060      0.000   7
age     0.058    0.003  0.019      0.876   1
height  0.042   -0.002  0.014      0.047   3
ankle   0.039    0.001  0.011      0.992  10
knee    0.036    0.001  0.015      0.963   9
chest   0.031    0.000  0.018      0.579   5
```

- PIP: 后验入选概率

BMA 系数估计及入选概率

```
> round(coef(mf, std.coefs = T, include.constant = F), 3)
      PIP Post Mean Post SD Cond.Pos.Sign Idx
abdomen 1.000      1.265  0.083          1.000  6
weight  0.965     -0.471  0.143          0.000  2
wrist   0.527     -0.079  0.085          0.000 13
forearm 0.306      0.034  0.057          1.000 12
neck    0.135     -0.017  0.050          0.000  4
biceps  0.124      0.013  0.041          1.000 11
thigh   0.094      0.011  0.041          0.999  8
hip     0.071     -0.013  0.060          0.000  7
age     0.058      0.003  0.019          0.876  1
height  0.042     -0.002  0.014          0.047  3
ankle   0.039      0.001  0.011          0.992 10
knee    0.036      0.001  0.015          0.963  9
chest   0.031      0.000  0.018          0.579  5
```

- PIP: 后验入选概率
- Post Mean 和 Post SD: 系数（效应）的估计值和标准差

BMA 系数估计及入选概率

```
> round(coef(mf, std.coefs = T, include.constant = F), 3)
      PIP Post Mean Post SD Cond.Pos.Sign Idx
abdomen 1.000    1.265  0.083      1.000   6
weight  0.965   -0.471  0.143      0.000   2
wrist   0.527   -0.079  0.085      0.000  13
forearm 0.306    0.034  0.057      1.000  12
neck    0.135   -0.017  0.050      0.000   4
biceps  0.124    0.013  0.041      1.000  11
thigh   0.094    0.011  0.041      0.999   8
hip     0.071   -0.013  0.060      0.000   7
age     0.058    0.003  0.019      0.876   1
height  0.042   -0.002  0.014      0.047   3
ankle   0.039    0.001  0.011      0.992  10
knee    0.036    0.001  0.015      0.963   9
chest   0.031    0.000  0.018      0.579   5
```

- PIP: 后验入选概率
- Post Mean 和 Post SD: 系数（效应）的估计值和标准差
- Cond.Pos.Sign: sign certainty, 入选条件下系数为正的后验概率

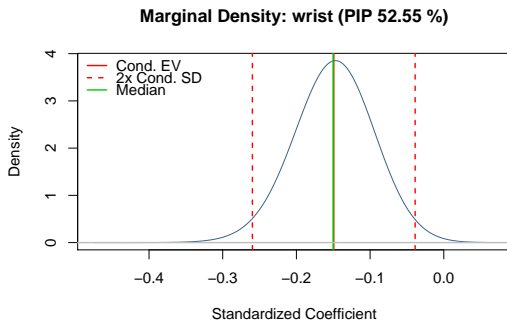
PMP 排名前 5 的模型

```
> topmodels.bma(mf)[, 1:5]
```

	0880	0881	0883	0885	0882
age	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
weight	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
height	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
neck	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
chest	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
abdomen	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
hip	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
thigh	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
knee	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
ankle	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
biceps	0.0000000	0.0000000	0.0000000	1.0000000	0.0000000
forearm	0.0000000	0.0000000	1.0000000	0.0000000	1.0000000
wrist	0.0000000	1.0000000	1.0000000	1.0000000	0.0000000
PMP (Exact)	0.2609014	0.1491782	0.1238701	0.03506753	0.02965686
PMP (MCMC)	0.2609014	0.1491782	0.1238701	0.03506753	0.02965686

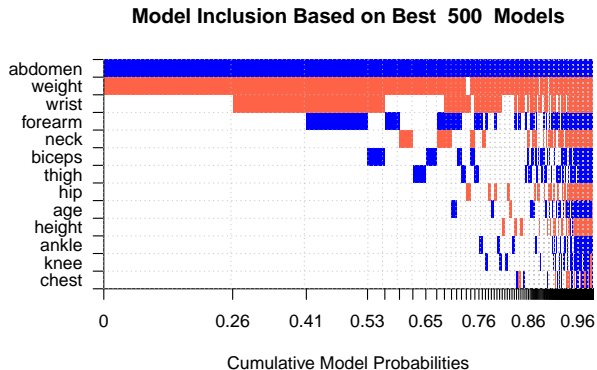
系数的后验密度 (Wrist)

```
density(mf,reg="wrist",std.coefs=TRUE)
```



所有入选模型图示

image(mf): 蓝色表示系数为正，红色为负，白色没入选。

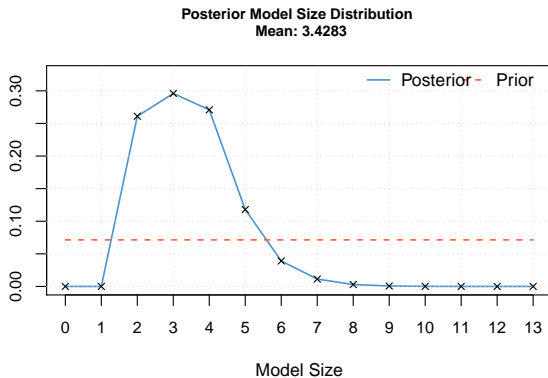


模型规模

`plotModelsize(mf)`

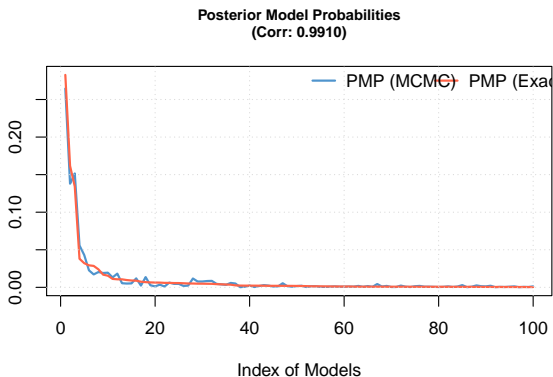
先验模型的平均规模为 $13/2 = 6.6$

数据修正后的后验模型平均规模为 3.43



比较 MCMC 模型与实际模型

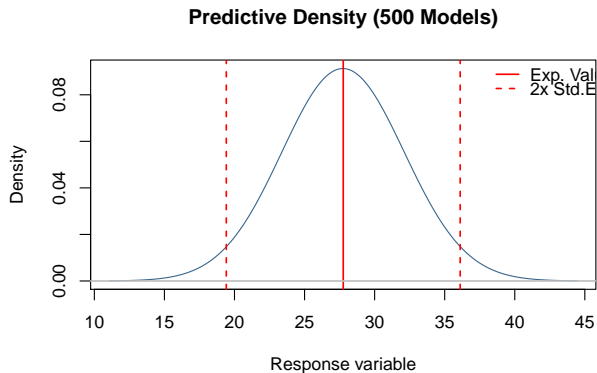
`plotConv(mf[1:100])` # 前 100 个模型



预测：用前 250 个数据预测最后一个数据

```
mf.p = bms(fat[1:250,], user.int = F)
dens.p = pred.density(mf.p, newdata = fat[251, ])
# 第 251 个人肥胖指数预测区间
quantile(dens.p, c(0.05, 0.95))
      5%      95%
20.58123 34.96956
# 第 251 个人肥胖指数预测分布
plot(dens.p)
```

第 251 个人肥胖指数的预测分布



预测表现：计算对数预测得分（LPS）

用前 200 个数据预测后 51 个人的肥胖指数，计算对数预测得分 (Log Predictive Score, LPS)

- 用于不同预测模型比较才有意义

```
mf.p = bms(fat[1:200, ], user.int = F)
lps.bma(mf.p,realized.y=fat[201:251,1],
        newdata=fat[201:251,-1])
[1] 2.962863
```

Outline

- 1 模型的不确定性
- 2 贝叶斯模型平均法基本原理
- 3 BMA 法在肥胖指数预测中的应用
- 4 BMA 软件与运用
- 5 讨论与结论**
- 6 参考文献

- ① 如何选择模型类型
 - ▶ 选择线性模型，对所有模型或部分模型加权平均
 - ▶ 找一个好模型，然后对该好模型所属类型加权平均
 - ▶ 对模型不同误差结构加权平均
- ② 模型平均的其它方法
 - ▶ 频率学派的 Bootstrap 方法
 - ▶ A minmax multiple shrinkage Stein estimator
 - ▶ 其它非贝叶斯方法：用 AIC、BIC、bootstrapping 等近似模型权重
 - ▶ 机器学习理论
- ③ 关于 BMA 在经济学模型中的最新应用，可参看：
STEEL M F J. Model Averaging and its Use in Economics[M]. arXiv preprint arXiv:1709.08221. 2017.

结论

- ① 在进行统计推断时，考虑模型的不确定性或统计结构的不确定性是非常重要的。
- ② 理论上，BMA 比任何单一模型的预测表现都好

批评意见：

- Too complicate to present easily
 - ▶ 可以仅考虑 PIP 做模型选择，比 p -值更容易理解
 - ▶ 可考虑 BMA 得到的单个“最佳”模型（视为考虑模型不确定性后得到的模型，或敏感性分析后的修正模型）
 - ▶ 避免了模型选择的不确定性（BMA 对模型选择比较稳健（Robust））
- Higher estimates of variance
 - ▶ BMA 才是正确的。
 - ▶ 单一模型无视模型的不确定性，低估了参数估计量的方差，从而过多地得出“显著”结论。

Outline

- 1 模型的不确定性
- 2 贝叶斯模型平均法基本原理
- 3 BMA 法在肥胖指数预测中的应用
- 4 BMA 软件与运用
- 5 讨论与结论
- 6 参考文献

- 1 HOETING J A, MADIGAN D, RAFTERY A E and Volinsky C T. Bayesian model averaging: a tutorial[J]. Statistical science, 1999: 382-401.
- 2 MORAL-BENITO E. Model averaging in economics: an overview[J]. Journal of Economic Surveys, 2015, 29(1): 46-75.
- 3 ZEUGNER S, FELDKIRCHER M. Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R[J]. Journal of Statistical Software, 2015, 68.
- 4 AMINI S M, PARMETER C F. Bayesian model averaging in R[J]. Journal of Economic and Social Measurement, 2011, 36(4): 253-287.
- 5 STEEL M F J. Model Averaging and its Use in Economics[M]. arXiv preprint arXiv:1709.08221. 2017.