第七章 多元线性回归贝叶斯模型

Wang Shujia

Department of Statistics, School of Economics
Shenzhen University

本章目标

1. 回顾传统多元线性回归模型

本章目标

1. 回顾传统多元线性回归模型
    - 估计

本章目标

1. 回顾传统多元线性回归模型
   - 估计
   - R 操作

本章目标

1. 回顾传统多元线性回归模型
    - 估计
    - R 操作
2. 建立贝叶斯多元线性回归模型

本章目标

1. 回顾传统多元线性回归模型
   - 估计
   - R 操作
2. 建立贝叶斯多元线性回归模型
   - 无信息先验

本章目标

1. 回顾传统多元线性回归模型
   - 估计
   - R 操作
2. 建立贝叶斯多元线性回归模型
   - 无信息先验
   - 共轭先验

本章目标

1. 回顾传统多元线性回归模型
    - 估计
    - R 操作
2. 建立贝叶斯多元线性回归模型
    - 无信息先验
    - 共轭先验
    - G-先验

本章目标

1. 回顾传统多元线性回归模型
    - 估计
    - R 操作
2. 建立贝叶斯多元线性回归模型
    - 无信息先验
    - 共轭先验
    - G-先验
3. 运用 WinBUGS 于回归模型

本章目标

1. 回顾传统多元线性回归模型
    - 估计
    - R 操作
2. 建立贝叶斯多元线性回归模型
    - 无信息先验
    - 共轭先验
    - G-先验
3. 运用 WinBUGS 于回归模型
    - 模型代码、抽样、收敛性诊断、结果

本章目标

1. 回顾传统多元线性回归模型
   - 估计
   - R 操作
2. 建立贝叶斯多元线性回归模型
   - 无信息先验
   - 共轭先验
   - G-先验
3. 运用 WinBUGS 于回归模型
   - 模型代码、抽样、收敛性诊断、结果
4. 运用 Rstan 于回归模型

本章目标

1. 回顾传统多元线性回归模型
   - 估计
   - R 操作
2. 建立贝叶斯多元线性回归模型
   - 无信息先验
   - 共轭先验
   - G-先验
3. 运用 WinBUGS 于回归模型
   - 模型代码、抽样、收敛性诊断、结果
4. 运用 Rstan 于回归模型
   1. 运用 Rstanarm 建立模型、MCMC 抽样、给出结果

本章目标

1. 回顾传统多元线性回归模型
   - 估计
   - R 操作
2. 建立贝叶斯多元线性回归模型
   - 无信息先验
   - 共轭先验
   - G-先验
3. 运用 WinBUGS 于回归模型
   - 模型代码、抽样、收敛性诊断、结果
4. 运用 Rstan 于回归模型
   1. 运用 Rstanarm 建立模型、MCMC 抽样、给出结果
   2. 运用 bayesplot 进行可视化

本章目标

1. 回顾传统多元线性回归模型
   - 估计
   - R 操作
2. 建立贝叶斯多元线性回归模型
   - 无信息先验
   - 共轭先验
   - G-先验
3. 运用 WinBUGS 于回归模型
   - 模型代码、抽样、收敛性诊断、结果
4. 运用 Rstan 于回归模型
   1. 运用 Rstanarm 建立模型、MCMC 抽样、给出结果
   2. 运用 bayesplot 进行可视化
   3. 先验分布的指定

Outline

1. 传统多元线性回归模型
2. 贝叶斯多元线性回归模型：理论
3. 贝叶斯多元线性回归模型：WinBUGS
4. 贝叶斯多元线性回归模型：RStan
5. 用 bayesplot 进行可视化
6. 先验分布的设定

Outline

1. 传统多元线性回归模型
   - 传统 MLR 模型：理论
   - 传统 MLR 模型：实例

多元线性回归模型 (MLR)

模型：

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i \ (i = 1, 2, \ldots, n)$$

- y_i 因变量，$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ 自变量
- $\boldsymbol{\beta} = (\beta_{1,}\beta_2, \ldots, \beta_p)^T$ 未知参数，通常 $x_{i1} = 1$，β_1 为截距

模型假设：

- $E[y_i|\boldsymbol{\beta}, \boldsymbol{x}_i] = \boldsymbol{x}_i^T \boldsymbol{\beta}$ （线性性）
- $Var[y_i|\boldsymbol{\beta}, \boldsymbol{x}_i] = \sigma^2$ （方差齐性）
- $y_i|\boldsymbol{\beta}, \boldsymbol{x}_i(i = 1, 2, \ldots, n)$ 相互独立（独立性）
- $\varepsilon_i \sim N(0, \sigma^2)$（正态性）

MLR 模型的矩阵表示

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

即

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times p}{\boldsymbol{X}}\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}},\ \boldsymbol{\varepsilon} \sim N_n(0, \sigma^2 \boldsymbol{I}_n)$$

最大似然估计与最小二乘估计

- 似然函数：

$$L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{n}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}$$

- 最大似然估计 (MLE)：使得 $L(L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, \boldsymbol{X})$ 最大，

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y},\ \hat{\sigma^2} = \frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n-p}$$

- 最小二乘估计 (LSE)：使得误差平方和最小

$$Q((\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

Outline

1. 传统多元线性回归模型
   - 传统 MLR 模型：理论
   - 传统 MLR 模型：实例

例 1：个人相貌会影响工资收入吗？

美国联邦调查局调查 1260 个工人的数据（beauty.csv）：

工资 (Wage) 时薪（单位：美元）

工会 (Union) 是否工会成员，是标示为 1，不是标示为 0

相貌 (Looks) 调查员给出的相貌吸引力评价，分五个档次 (1=homely, 2=quite plain, 3=average, 4=good looking, 5= strikingly beautiful or handsome)

教育程度 (Educ) 以年限来表示受教育程度

经验 (Exper) 潜在的工作经验 (以年限表示)，定义为年龄减去受教育年限减去 6 (假定学校教育从 6 岁开始)

女性 (Female) 性别，女性标示为 1，男性标示为 0

健康状况 (Goodhlth) 1= 健康，0= 不健康

婚姻 (Married) 1= 已婚，0= 未婚

城市规模 (Bigcity) 1= 大城市，0= 小城市

种族 (Black) 1= 黑人，0= 否则

在 R 中运行线性回归模型

```
beauty <- read_csv("F:/Teaching/Rdata/beauty.csv")
out_looks <- lm(wage ~ looks, data = beauty)
summary(out_looks)
```

```
Call:
lm(formula = wage ~ looks, data = beauty)

Residuals:
   Min     1Q Median     3Q    Max
-5.452 -2.737 -0.997  1.453 71.108

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.1139     0.6242   8.192 6.24e-16 ***
looks         0.3744     0.1916   1.954   0.0509 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.655 on 1258 degrees of freedom
Multiple R-squared:  0.003027,	Adjusted R-squared:  0.002235
F-statistic:  3.82 on 1 and 1258 DF,  p-value: 0.05088
```

定性变量转换为虚拟变量

```
beauty <- beauty  %>%
mutate(looks_dum = as.factor(looks))
out_looks <- lm(wage ~ looks_dum, data = beauty)
summary(out_looks)
```

```
Call:
lm(formula = wage ~ looks_dum, data = beauty)

Residuals:
   Min     1Q Median     3Q    Max
-5.485 -2.700 -0.950  1.401 71.421

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.6215     1.2894   3.584 0.000351 ***
looks_dum2    0.7073     1.3471   0.525 0.599663
looks_dum3    1.8831     1.3009   1.447 0.148020
looks_dum4    1.6778     1.3122   1.279 0.201273
looks_dum5    2.7669     1.6733   1.654 0.098477 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.649 on 1255 degrees of freedom
Multiple R-squared:  0.008163,  Adjusted R-squared:  0.005002
F-statistic: 2.582 on 4 and 1255 DF,  p-value: 0.03574
```

个人相貌真的对工资收入无显著影响吗？

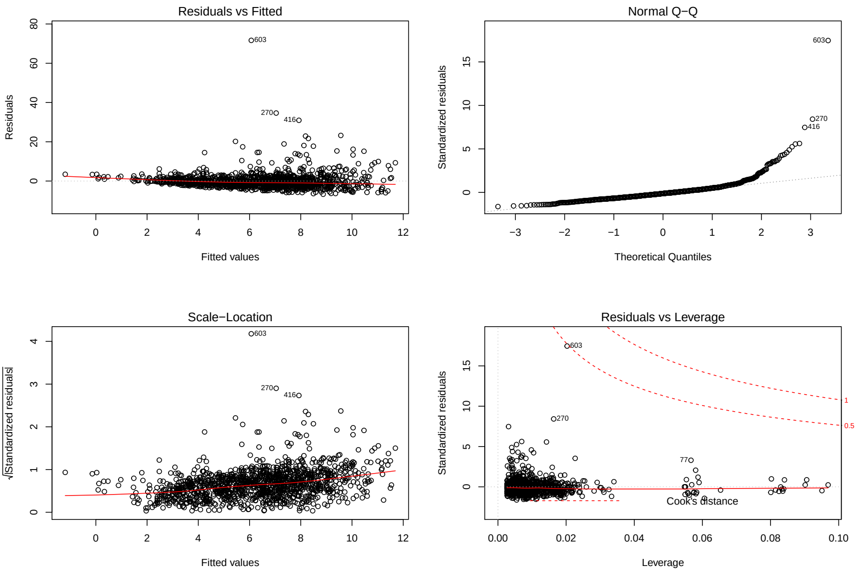
```
Call:
lm(formula = wage ~ educ + exper + looks_dum + union + goodhlth +
    black + female + married + bigcity, data = beauty)

Residuals:
   Min     1Q Median     3Q    Max
-6.734 -2.100 -0.531  1.146 71.653

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.74556    1.37754  -1.267  0.20534
educ         0.41725    0.04721   8.839  < 2e-16 ***
exper        0.07652    0.01068   7.162 1.35e-12 ***
looks_dum2   0.50555    1.20934   0.418  0.67599
looks_dum3   1.28107    1.16909   1.096  0.27338
looks_dum4   1.35854    1.18415   1.147  0.25149
looks_dum5   2.79617    1.50702   1.855  0.06377 .
union        0.65351    0.26745   2.443  0.01468 *
goodhlth    -0.02916    0.47602  -0.061  0.95116
black       -0.15918    0.46238  -0.344  0.73071
female      -2.26688    0.26638  -8.510  < 2e-16 ***
married      0.77785    0.27492   2.829  0.00474 **
bigcity      1.36503    0.29063   4.697 2.94e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.147 on 1247 degrees of freedom
Multiple R-squared:  0.2159,    Adjusted R-squared:  0.2084
F-statistic: 28.62 on 12 and 1247 DF,  p-value: < 2.2e-16
```

诊断图：4 in 1


Residuals vs Fitted
Residuals
Fitted values
603
270
416
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Cook's distance
77
1
0.5

```
hist(beauty$wage, xlab="wage", probability = T,
     main = "Histogram of wage")
```

对 wage 做对数变换

```
hist(log(beauty$wage), xlab="wage", probability = T,
     main = "Histogram of log(wage)")
```

```
out_log <- lm(log(wage) ~ educ + exper + looks_dum
              + union + goodhlth + black + female
              + married + bigcity, data = beauty)
```

```
summary(out_log)
```

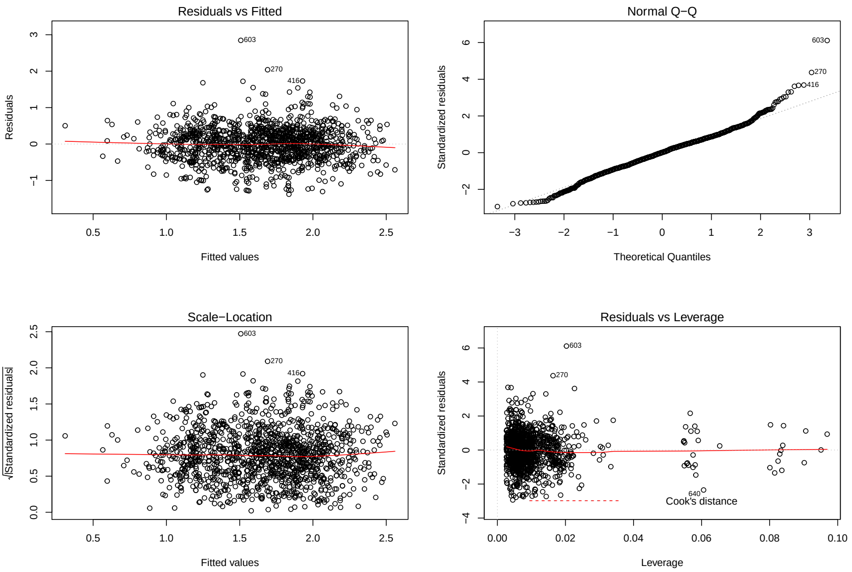
```
Call:
lm(formula = log(wage) ~ educ + exper + look_dum + union + goodhlth +
    black + female + married + bigcity, data = beauty)

Residuals:
     Min       1Q   Median       3Q      Max
-1.37855 -0.30460  0.00564  0.28476  2.84543

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.276382   0.156275   1.769   0.0772 .
educ         0.067074   0.005355  12.525  < 2e-16 ***
exper        0.012860   0.001212  10.610  < 2e-16 ***
look_dum2    0.150341   0.137193   1.096   0.2734
look_dum3    0.275556   0.132627   2.078   0.0379 *
look_dum4    0.275103   0.134335   2.048   0.0408 *
look_dum5    0.420596   0.170963   2.460   0.0140 *
union        0.182004   0.030341   5.999 2.60e-09 ***
goodhlth     0.070962   0.054001   1.314   0.1891
black       -0.108573   0.052454  -2.070   0.0387 *
female      -0.419692   0.030220 -13.888  < 2e-16 ***
married      0.061428   0.031188   1.970   0.0491 *
bigcity      0.182370   0.032971   5.531 3.87e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4704 on 1247 degrees of freedom
Multiple R-squared:  0.3799,    Adjusted R-squared:  0.3739
F-statistic: 63.65 on 12 and 1247 DF,  p-value: < 2.2e-16
```

诊断图 (out_log)：4 in 1



Outline

1 传统多元线性回归模型

2 贝叶斯多元线性回归模型：理论

3 贝叶斯多元线性回归模型：WinBUGS

4 贝叶斯多元线性回归模型：RStan

5 用 bayesplot 进行可视化

6 先验分布的设定

Outline

2 贝叶斯多元线性回归模型：理论
- 贝叶斯 MLR 模型
- 无信息先验
- 共轭先验
- Zellner G-先验

贝叶斯多元线性回归模型

（一）给出解释变量及其分布：

$$y_i \sim N(\mu_i, \sigma^2),\ (i = 1, \ldots, n)$$

其中

$$\mu_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p$$

或

$$\boldsymbol{y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$$

（二）给出未知参数 $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)^T$ 的先验分布：

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)^T \sim \pi(\boldsymbol{\theta}),$$

（三）后验分布为

$$p(\boldsymbol{\theta}|\boldsymbol{X}) \propto \pi(\boldsymbol{\theta}) L(\boldsymbol{\theta})$$

什么情况下用贝叶斯 MLR 模型？

1. 下面情形传统多元线性回归 OK：
   - 观察数据足够多
   - 模型假设完全满足
2. 下面情形贝叶斯方法更合适：
   1. 数据中等或偏少。可以用先验信息提高推断的准确性
   2. 模型假设不成立（独立性、方差齐性、正态性）。贝叶斯方法简单直接
   3. 总体非正态分布。贝叶斯方法可以直接设置（如厚尾分布，改善模型稳健性）
   4. 复杂模型。贝叶斯方法结合 MCMC 算法是好工具（如 DSGE 模型）

Outline

2 贝叶斯多元线性回归模型：理论

- 贝叶斯 MLR 模型
- 无信息先验
- 共轭先验
- Zellner G-先验

无信息先验

模型：

$$\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I}_n)$$

先验：Jeffreys 无信息先验

$$(\boldsymbol{\beta},\sigma^2) \propto \sigma^{-2}$$

或

$$\boldsymbol{\beta} \propto 1\ (-\infty < \beta < +\infty),\ \sigma^2 \propto 1/\sigma^2\ (\sigma > 0)$$

实践中注意：

1. WinBUGS 中的**正态分布**给出的是精度 $\tau = 1/\sigma^2$，而不是方差 σ^2

无信息先验

模型：

$$\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I}_n)$$

先验：Jeffreys 无信息先验

$$(\boldsymbol{\beta},\sigma^2) \propto \sigma^{-2}$$

或

$$\boldsymbol{\beta} \propto 1\ (-\infty < \beta < +\infty),\ \sigma^2 \propto 1/\sigma^2\ (\sigma > 0)$$

实践中注意：

1. WinBUGS 中的**正态分布**给出的是精度 $\tau = 1/\sigma^2$，而不是方差 σ^2
2. Stan 中的**正态分布**给出的是标准差 σ

无信息先验

模型：

$$\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$$

先验：Jeffreys 无信息先验

$$(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}$$

或

$$\boldsymbol{\beta} \propto 1 \ (-\infty < \beta < +\infty), \ \sigma^2 \propto 1/\sigma^2 \ (\sigma > 0)$$

实践中注意：

1. WinBUGS 中的**正态分布**给出的是精度 $\tau = 1/\sigma^2$，而不是方差 σ^2
2. Stan 中的**正态分布**给出的是标准差 σ
3. 无信息先验的指定：

无信息先验

模型：

$$\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I}_n)$$

先验：Jeffreys 无信息先验

$$(\boldsymbol{\beta},\sigma^2) \propto \sigma^{-2}$$

或

$$\boldsymbol{\beta} \propto 1 \ (-\infty < \beta < +\infty),\ \sigma^2 \propto 1/\sigma^2 \ (\sigma > 0)$$

实践中注意：

1. WinBUGS 中的**正态分布**给出的是精度 $\tau = 1/\sigma^2$，而不是方差 σ^2
2. Stan 中的**正态分布**给出的是标准差 σ
3. 无信息先验的指定：
    1. 常用 $\beta_i \sim N(0,10^{-3})$ 或 $\beta_i \sim dflat()$

无信息先验

模型：

$$\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I}_n)$$

先验：Jeffreys 无信息先验

$$(\boldsymbol{\beta},\sigma^2) \propto \sigma^{-2}$$

或

$$\boldsymbol{\beta} \propto 1\ (-\infty < \beta < +\infty),\ \sigma^2 \propto 1/\sigma^2\ (\sigma > 0)$$

实践中注意：

1. WinBUGS 中的**正态分布**给出的是精度 $\tau = 1/\sigma^2$，而不是方差 σ^2
2. Stan 中的**正态分布**给出的是标准差 σ
3. 无信息先验的指定：
    1. 常用 $\beta_i \sim N(0, 10^{-3})$ 或 $\beta_i \sim dflat()$
    2. 精度的先验：$\tau = 1/\sigma^2 \sim Gamma(10^{-3}, 10^{-3})$，或 $\sigma^2 \sim IG(10^3, 10^3)$

无信息先验

模型：

$$\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I}_n)$$

先验：Jeffreys 无信息先验

$$(\boldsymbol{\beta},\sigma^2) \propto \sigma^{-2}$$

或

$$\boldsymbol{\beta} \propto 1\ (-\infty < \beta < +\infty),\ \sigma^2 \propto 1/\sigma^2\ (\sigma > 0)$$

实践中注意：

1. WinBUGS 中的**正态分布**给出的是精度 $\tau = 1/\sigma^2$，而不是方差 σ^2
2. Stan 中的**正态分布**给出的是标准差 σ
3. 无信息先验的指定：
   1. 常用 $\beta_i \sim N(0, 10^{-3})$ 或 $\beta_i \sim dflat()$
   2. 精度的先验：$\tau = 1/\sigma^2 \sim Gamma(10^{-3}, 10^{-3})$，或 $\sigma^2 \sim IG(10^3, 10^3)$
   3. $\log(\sigma^2) = 2\log\sigma, \log\sigma \sim dflat()$

后验分布

- 条件后验分布：$\boldsymbol{\beta}|\sigma^2, \boldsymbol{y}, \boldsymbol{X} \sim N(\hat{\boldsymbol{\beta}}, \boldsymbol{V}_\beta \sigma^2)$，其中

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$
$$\boldsymbol{V}_\beta = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}$$

- 边缘后验分布：$\sigma^2|\boldsymbol{y}, \boldsymbol{X} \sim \mathrm{IG}((n-p)/2, (n-p)s^2/2)$，其中

$$s^2 = \frac{1}{n-p}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

- 边缘后验分布：

$$\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X} \sim t_{n-p}(\hat{\boldsymbol{\beta}}, \frac{(n-p)s^2}{n-p-2}\boldsymbol{V}_\beta),$$

  此分布在实际计算中一般不用。
- 联合后验分布：$(\boldsymbol{\beta}, \sigma^2)|(\boldsymbol{y}, \boldsymbol{X}) \sim N(\hat{\boldsymbol{\beta}}, \boldsymbol{V}_\beta\sigma^2) \times \mathrm{Inv}-\chi^2(n-p, s^2)$

预测分布

- 假设我们得到自变量新的数据 $\tilde{\boldsymbol{X}}$，需要预测因变量的值 $\tilde{\boldsymbol{y}}$
  - 假如 $(\boldsymbol{\beta}, \sigma^2)$ 已知，则有 $\tilde{\boldsymbol{y}}|\boldsymbol{\beta}, \sigma^2 \sim N(\tilde{\boldsymbol{X}}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$
  - 当然，现在对 $(\boldsymbol{\beta}, \sigma^2)$ 的了解已经基于它的后验分布
- 后验预测分布：$p(\tilde{\boldsymbol{y}}|\boldsymbol{y})$ 为

$$\tilde{\boldsymbol{y}}|\boldsymbol{y}, \boldsymbol{X} \sim t_{n-p}[\tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}, (\boldsymbol{I} + \tilde{\boldsymbol{X}}\boldsymbol{V}_\beta \tilde{\boldsymbol{X}}^T s^2)]$$

- MCMC 抽样：得到 $(\boldsymbol{\beta}^{(j)}, \sigma^{2(j)})$ 后，$\tilde{\boldsymbol{y}}^{(j)} \sim N(\tilde{\boldsymbol{X}}\boldsymbol{\beta}^{(j)}, \sigma^{2(j)}\boldsymbol{I}_n)$ 即为后验预测分布的样本。

Outline

2. 贝叶斯多元线性回归模型：理论
   - 贝叶斯 MLR 模型
   - 无信息先验
   - 共轭先验
   - Zellner G-先验

共轭先验

模型：

$$\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2 \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$$

共轭先验：

$$\boldsymbol{\beta}|\sigma^2 \sim N(\mathbb{B}, \sigma^2 \boldsymbol{V})$$

$$\sigma^2 \sim \mathrm{IG}(a, b), (a, b > 0)$$

超参数的含义：

1. $\mathbb{B}$ 是 $\boldsymbol{\beta}$ 的先验均值，通常取值 0
2. $\sigma^2\boldsymbol{V}$ 为 $\boldsymbol{\beta}$ 的先验协方差矩阵（正定），常取 $\boldsymbol{V} = c^{-1}I$，c 越大，先验越集中于 $\mathbb{B}$。也可取 $\boldsymbol{V}^{-1} = \mathrm{diag}(c_1, c_2, \ldots, c_p)$
3. a, b 为 σ^2 的先验分布参数（预先给定）

后验分布

1. 联合后验分布：

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) &\propto \pi(\boldsymbol{\beta}, \sigma^2) N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n) \\ &\propto \mathrm{IG}(a, b) N(\mathbb{B}, \sigma^2 \boldsymbol{V}) N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n) \end{aligned}$$

2. σ^2 的边缘后验分布：

$$\sigma^2 | \boldsymbol{y}, \boldsymbol{X} \sim \mathrm{IG}\left(\frac{n}{2} + a, b + \frac{(n-p)s^2}{2} + \frac{s^*}{2}\right)$$

其中

$$s^* = (\mathbb{B} - \hat{\boldsymbol{\beta}})^T [\boldsymbol{V} + (\boldsymbol{X}^T \boldsymbol{X})^{-1}]^{-1} (\mathbb{B} - \hat{\boldsymbol{\beta}})$$

3. $\boldsymbol{\beta}$ 的条件后验分布：

$$\boldsymbol{\beta} | \sigma^2, \boldsymbol{y}, \boldsymbol{X} \sim N_p\left(\boldsymbol{\mu}_p, \sigma^2 (\boldsymbol{V}^{-1} + \boldsymbol{X}^T \boldsymbol{X})^{-1}\right)$$

其中

$$\boldsymbol{\mu}_p = (\boldsymbol{V}^{-1} + \boldsymbol{X}^T \boldsymbol{X})^{-1} [(\boldsymbol{X}^T \boldsymbol{X}) \hat{\boldsymbol{\beta}} + \boldsymbol{V}^{-1} \mathbb{B}]$$

贝叶斯点估计

$$
\begin{aligned}
E(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}) &= \boldsymbol{\mu}_p \\
E(\sigma^2|\boldsymbol{y}, \boldsymbol{X}) &= \frac{1}{(n+2a-2)}\left(2b + (n-p)s^2 + s^*\right)
\end{aligned}
$$

Outline

2 贝叶斯多元线性回归模型：理论

- 贝叶斯 MLR 模型
- 无信息先验
- 共轭先验
- Zellner G-先验

Zellner G-先验

预先给定回归系数的先验均值 $\mathbb{B}$ 和先验强度参数 c，

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{X} \sim N(\mathbb{B}, c\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$$

$$\sigma^2|\boldsymbol{X} \sim \pi(\sigma^2|\boldsymbol{X}) \propto \sigma^{-2}$$

其中 $1/c$ 可解释为先验信息与数据信息的比例：$c = 1$ 表示相同权重；$1/c = 0.5$ 表示先验信息和数据权重各占 50%。

后验分布

1. σ^2 的边缘后验分布：

$$\sigma^2|\boldsymbol{y},\boldsymbol{X} \sim \mathrm{IG}\left(\frac{n}{2}, \frac{(n-p)s^2}{2} + \frac{1}{2(c+1)}(\mathbb{B}-\hat{\boldsymbol{\beta}})^T(\boldsymbol{X}^T\boldsymbol{X})(\mathbb{B}-\hat{\boldsymbol{\beta}})\right)$$

2. $\boldsymbol{\beta}$ 的条件后验分布：

$$\boldsymbol{\beta}|\sigma^2,\boldsymbol{y},\boldsymbol{X} \sim N_p\left(\frac{c}{c+1}(\mathbb{B}/c+\hat{\boldsymbol{\beta}}), \frac{c\sigma^2}{c+1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)$$

贝叶斯点估计

$$E(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}) = \frac{1}{c+1}(\mathbb{B} + c\hat{\boldsymbol{\beta}})$$
$$E(\sigma^2|\boldsymbol{y}, \boldsymbol{X}) = \frac{1}{(n-2)}\left((n-p)s^2 + \frac{1}{c+1}(\mathbb{B} - \hat{\boldsymbol{\beta}})^T(\boldsymbol{X}^T\boldsymbol{X})(\mathbb{B} - \hat{\boldsymbol{\beta}})\right)$$

显然，当 $c \to \infty$ 时，先验信息对后验估计的影响消失了。

Outline

1. 传统多元线性回归模型
2. 贝叶斯多元线性回归模型：理论
3. 贝叶斯多元线性回归模型：WinBUGS
4. 贝叶斯多元线性回归模型：RStan
5. 用 bayesplot 进行可视化
6. 先验分布的设定

贝叶斯线性回归模型：无信息先验

数据模型：

$$\log(wage)_i \sim N(\mu_i, \sigma^2)$$
$$\mu_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 female_i + \beta_4 married_i$$

无信息先验分布：
$\beta_i \sim dflat(), i = 0, 1, 2, 3, 4$
$\sigma^2 \sim IG(10^3, 10^3)$ 或 $\tau = 1/\sigma^2 \sim Gamma(0.001, 0.001)$

运用 R2WinBUGS 步骤

1. 准备模型代码
2. 准备数据（list 数据结构）
3. 指定参数和初始值
4. 调用 R2WinBUGS，运行 bugs 函数 (包括数据、初始值、参数，迭代次数等等）
5. MCMC 收敛性判断
6. 结果解释

第一步：模型代码（保存为.bug 文件）

```
model{
#Likelihood:
for(i in 1:N){
   y[i]~dnorm(mu[i],tau)
   mu[i]<-beta0+inprod(beta[1:np],x[i,1:np])
   }
#Prior
tau~dgamma(0.001,0.001)
sigma2<-1/sqrt(tau)
beta0~dflat()
for ( j in 1:np ){
     beta[j] ~ dflat()
     }
}
```

第二步：准备数据（list 结构）

```
beauty <- read_csv("F:/Rdata/beauty.csv")
#Specify independent var and response var
  xdata <- beauty %>%
  select("educ","exper","female","married")
y <- log(beauty$wage)
x <- as.matrix(xdata)# x must be a matrix
N <- nrow(x)
np <- ncol(x)
data <- list(y = y, x = x, N = N, np = np)
```

第三步：指定参数和初始值

```r
#Specify parameters
parameters <- list("beta0","beta","tau")
#Prepare initials
inits <- function(){
  list(beta0=rnorm(1), beta=rnorm(np,0,10), tau=runif(1,0,10))
}
```

第四步：运行 R2WinBUGS

```
output<-bugs(
  data,
  inits,
  parameters,
  n.chains=3,
  n.iter=2000,
  n.burnin=1000,
  n.thin=1,
  debug=FALSE,
  codaPkg=FALSE,
  model.file="F:\\Rdata\\Ch7_MLRmodel.bug",
  bugs.directory="C:\\WinBUGS14\\"
  )
```

第五步：判断 MCMC 收敛性

MCMC 收敛性判断：一般只能判断哪种情况不收敛，不能证明其收敛。

1. Trace 或 History：是否平稳；多条链是否重合

第五步：判断 MCMC 收敛性

MCMC 收敛性判断：一般只能判断哪种情况不收敛，不能证明其收敛。

1. Trace 或 History：是否平稳；多条链是否重合
2. 后验分布的密度函数：光滑程度

第五步：判断 MCMC 收敛性

MCMC 收敛性判断：一般只能判断哪种情况不收敛，不能证明其收敛。

1. Trace 或 History：是否平稳；多条链是否重合
2. 后验分布的密度函数：光滑程度
3. 自相关图：是否存在自相关

第五步：判断 MCMC 收敛性

MCMC 收敛性判断：一般只能判断哪种情况不收敛，不能证明其收敛。

1. Trace 或 History：是否平稳；多条链是否重合
2. 后验分布的密度函数：光滑程度
3. 自相关图：是否存在自相关
4. R-hat：缩减因子，等于 1 表示收敛

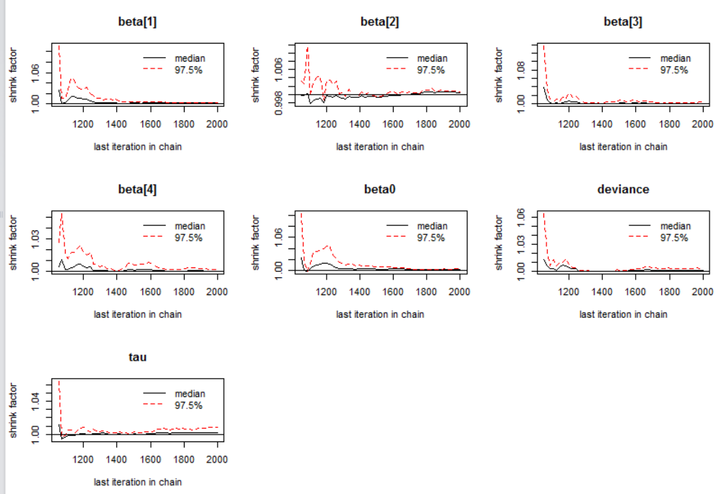
第五步：判断 MCMC 收敛性

MCMC 收敛性判断：一般只能判断哪种情况不收敛，不能证明其收敛。

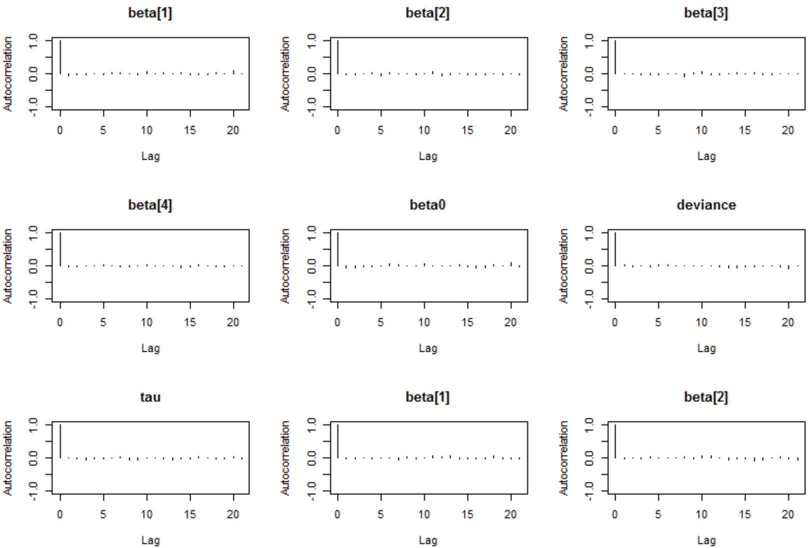
1. Trace 或 History：是否平稳；多条链是否重合
2. 后验分布的密度函数：光滑程度
3. 自相关图：是否存在自相关
4. R-hat：缩减因子，等于 1 表示收敛
5. n.eff: 有效样本容量 (effective sample size)

gelman.plot()

```
A<-as.mcmc.list(output)
gelman.plot(A)
```



自相关：autocorr.plot()

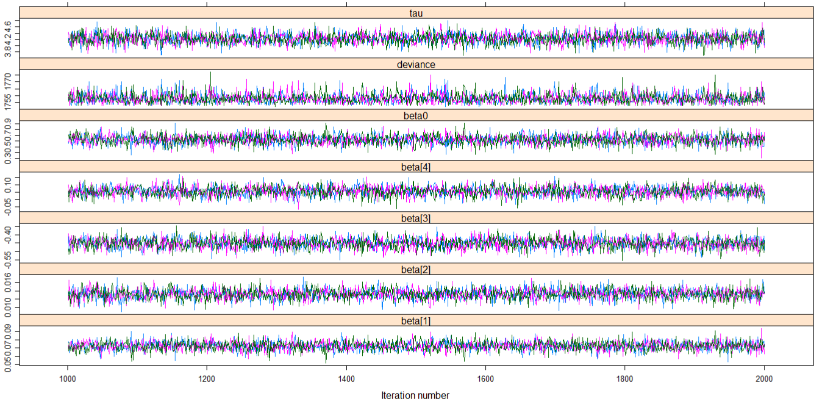


自相关：autocorr.diag()

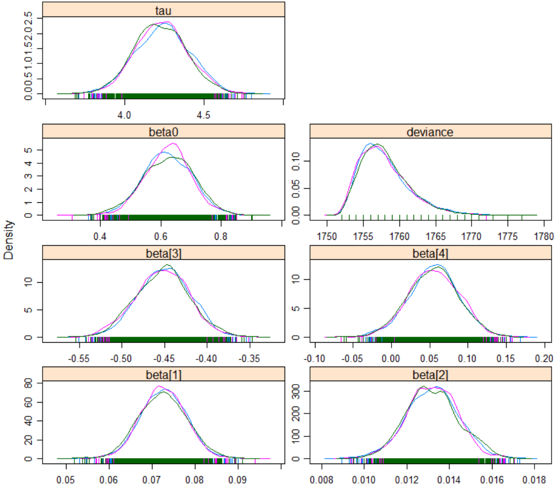
```
> autocorr.diag(A)
           beta[1]      beta[2]      beta[3]      beta[4]
Lag 0   1.00000000  1.000000000  1.000000000  1.000000000
Lag 1  -0.02639854 -0.004158444  0.005137248  0.010585820
Lag 5   0.01004456 -0.017628321  0.007743063  0.006710833
Lag 10  0.01561493  0.019258180  0.015090827 -0.007931777
Lag 50  0.02589334 -0.014220770 -0.016228800  0.016969987
             beta0     deviance         tau
Lag 0   1.00000000  1.000000000  1.00000000
Lag 1  -0.02684240 -0.004350204 -0.01002919
Lag 5   0.01901889  0.023543968 -0.01626173
Lag 10  0.01545915 -0.005944197  0.01346949
Lag 50  0.01605285  0.023110788 -0.01473964
```

Traceplot

```
library(lattice)
xyplot(A)
```



densityplot(A)



第六步：结果解释

```
> print(output,digit=3)
Inference for Bugs model at "F:\BaiduYun\Teaching\Rdata\Ch5model.txt", fit using WinBUGS,
 3 chains, each with 10000 iterations (first 3000 discarded)
 n.sims = 21000 iterations saved
             mean    sd     2.5%      25%      50%      75%    97.5%  Rhat n.eff
beta0        0.621 0.080    0.467    0.567    0.621    0.676    0.777 1.001  6500
beta[1]      0.073 0.005    0.062    0.069    0.073    0.076    0.083 1.001 11000
beta[2]      0.013 0.001    0.011    0.012    0.013    0.014    0.016 1.001 21000
beta[3]     -0.451 0.031   -0.513   -0.472   -0.450   -0.430   -0.392 1.001 18000
beta[4]      0.056 0.032   -0.006    0.034    0.056    0.077    0.118 1.001 21000
tau          4.234 0.171    3.906    4.116    4.231    4.348    4.578 1.001 21000
deviance  1758.065 3.526 1753.000 1755.000 1757.000 1760.000 1767.000 1.001 10000

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = Dbar-Dhat)
pD = 6.0 and DIC = 1764.1
DIC is an estimate of expected predictive error (lower deviance is better).
```

1. 哪些解释变量对工资收入有显著影响？

第六步：结果解释

```
> print(output,digit=3)
Inference for Bugs model at "F:\BaiduYun\Teaching\Rdata\ch5model.txt", fit using WinBUGS,
 3 chains, each with 10000 iterations (first 3000 discarded)
 n.sims = 21000 iterations saved
             mean    sd     2.5%      25%      50%      75%     97.5%  Rhat n.eff
beta0       0.621 0.080    0.467    0.567    0.621    0.676    0.777 1.001  6500
beta[1]     0.073 0.005    0.062    0.069    0.073    0.076    0.083 1.001 11000
beta[2]     0.013 0.001    0.011    0.012    0.013    0.014    0.016 1.001 21000
beta[3]    -0.451 0.031   -0.513   -0.472   -0.450   -0.430   -0.392 1.001 18000
beta[4]     0.056 0.032   -0.006    0.034    0.056    0.077    0.118 1.001 21000
tau         4.234 0.171    3.906    4.116    4.231    4.348    4.578 1.001 21000
deviance 1758.065 3.526 1753.000 1755.000 1757.000 1760.000 1767.000 1.001 10000

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = Dbar-Dhat)
pD = 6.0 and DIC = 1764.1
DIC is an estimate of expected predictive error (lower deviance is better).
```

1. 哪些解释变量对工资收入有显著影响？
2. 存在性别歧视吗？

第六步：结果解释

```
> print(output,digit=3)
Inference for Bugs model at "F:\BaiduYun\Teaching\Rdata\ch5model.txt", fit using WinBUGS,
 3 chains, each with 10000 iterations (first 3000 discarded)
 n.sims = 21000 iterations saved
             mean    sd     2.5%      25%      50%      75%     97.5%  Rhat n.eff
beta0       0.621 0.080    0.467    0.567    0.621    0.676    0.777 1.001  6500
beta[1]     0.073 0.005    0.062    0.069    0.073    0.076    0.083 1.001 11000
beta[2]     0.013 0.001    0.011    0.012    0.013    0.014    0.016 1.001 21000
beta[3]    -0.451 0.031   -0.513   -0.472   -0.450   -0.430   -0.392 1.001 18000
beta[4]     0.056 0.032   -0.006    0.034    0.056    0.077    0.118 1.001 21000
tau         4.234 0.171    3.906    4.116    4.231    4.348    4.578 1.001 21000
deviance 1758.065 3.526 1753.000 1755.000 1757.000 1760.000 1767.000 1.001 10000

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = Dbar-Dhat)
pD = 6.0 and DIC = 1764.1
DIC is an estimate of expected predictive error (lower deviance is better).
```

1. 哪些解释变量对工资收入有显著影响？
2. 存在性别歧视吗？
3. 结婚对工资收入有影响吗？

第六步：结果解释

```
> print(output,digit=3)
Inference for Bugs model at "F:\BaiduYun\Teaching\Rdata\ch5model.txt", fit using WinBUGS,
 3 chains, each with 10000 iterations (first 3000 discarded)
 n.sims = 21000 iterations saved
             mean    sd     2.5%     25%      50%      75%    97.5%  Rhat n.eff
beta0       0.621 0.080    0.467    0.567    0.621    0.676    0.777 1.001  6500
beta[1]     0.073 0.005    0.062    0.069    0.073    0.076    0.083 1.001 11000
beta[2]     0.013 0.001    0.011    0.012    0.013    0.014    0.016 1.001 21000
beta[3]    -0.451 0.031   -0.513   -0.472   -0.450   -0.430   -0.392 1.001 18000
beta[4]     0.056 0.032   -0.006    0.034    0.056    0.077    0.118 1.001 21000
tau         4.234 0.171    3.906    4.116    4.231    4.348    4.578 1.001 21000
deviance 1758.065 3.526 1753.000 1755.000 1757.000 1760.000 1767.000 1.001 10000

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = Dbar-Dhat)
pD = 6.0 and DIC = 1764.1
DIC is an estimate of expected predictive error (lower deviance is better).
```

1. 哪些解释变量对工资收入有显著影响？
2. 存在性别歧视吗？
3. 结婚对工资收入有影响吗？
4. 受教育年限对工资收入有影响吗？

Outline

1. 传统多元线性回归模型
2. 贝叶斯多元线性回归模型：理论
3. 贝叶斯多元线性回归模型：WinBUGS
4. 贝叶斯多元线性回归模型：RStan
5. 用 bayesplot 进行可视化
6. 先验分布的设定

Outline

4. 贝叶斯多元线性回归模型：RStan
   - 软件包介绍
   - 用 rstanarm 运行贝叶斯模型

贝叶斯模型软件包

对于标准模型，首选软件包是 rstanarm 和 brms，而不是直接 RStan。

- rstanarm: Bayesian Applied Regression Modeling via Stan
  - 预先编译，速度快
- brms: R package for Bayesian generalized multivariate non-linear multilevel models using Stan
  - 模型代码需要编译，速度慢
  - 扩展性强，更多模型和功能
- 两个包都可以解决常用模型
  - Standard Regression and GLM
  - Categorical Models
  - Mixed Models
- 都可以用 bayesplot 于模型可视化
- 都可以用 loo 于模型的选择与比较

rstanarm

- Standard Regression and GLM
  - stan_aov: ANOVA
  - stan_lm: standard regression (不推荐，因为需要 R^2 先验)
  - stan_glm: generalized linear model (推荐，可用 bayes_R2 获取 R^2)
  - stan_glm.nb: negative binomial for count data or neg_binomial_2 family for stan_glm
  - stan_polr: ordinal regression model
  - stan_biglm: big data lm
- Mixed Models
  - stan_lmer: standard lme4 style mixed model
  - stan_glmer: glmm(generalized linear mixed model)
  - stan_glmer.nb: for negative binomial
  - stan_nlmer: nlme, Bayesian nonlinear models with group-specific terms
  - stan_mvmer: multivariate outcome
  - stan_gamm4: generalized additive mixed model in lme4 style

Outline

4. 贝叶斯多元线性回归模型：RStan
   - 软件包介绍
   - 用 rstanarm 运行贝叶斯模型

RStanArm 能做什么？

The goal of the *rstanarm* package is to make Bayesian estimation *routine* for the most common regression models that applied researchers use. This will enable researchers to avoid the counter-intuitiveness of the frequentist approach to probability and statistics with only minimal changes to their existing R scripts.

- models are specified with formula syntax,
- data is provided as a data frame, and
- additional arguments are available to specify priors.

Estimation may be carried out with Markov chain Monte Carlo, variational inference, or optimization (Laplace approximation). Graphical posterior predictive checking, leave-one-out cross-validation, and posterior visualization are tightly integrated.

例 1：个人相貌会影响工资收入吗？

数据模型：

$$\log(wage)_i \sim N(\mu_i, \sigma^2)(i = 1, \ldots, n)$$
$$\mu_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 female_i + \beta_4 married_i$$

先验：各参数相互独立
后验：

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \Pi_{k=1}^{4} \pi(\beta_k)\pi(\sigma^2)\mathrm{Normal}(\mu_i, \sigma^2)$$

准备数据

```
#Load data set
beauty <- read_csv("F:/Rdata/beauty.csv")
#Specify independent var and response var
  wage_data <- beauty %>%
    mutate(lwage = log(wage)) %>%
    select("lwage","educ","exper","female","married")
```

运行 rstanarm

```
library(rstanarm)
post <- stan_glm(
          lwage ~ educ + exper + female + married,
          data = wage_data,
          family = gaussian(link = "identity"),
          chains = 2,cores=2,seed=123456,iter = 500)
```

默认：chains=4,iter=2000，前面 1000 舍去

输出结果：print(post,digits = 3)

给出参数估计的 Median 和 MAD_SD，
 其中 Median Absolute Deviation(MAD) 是后验标准差的稳健估计。

运行结果

```
stan_glm
 family:       gaussian [identity]
 formula:      lwage ~ educ + exper + female + married
 observations: 1260
 predictors:   5
------
            Median MAD_SD
(Intercept)  0.630  0.075
educ         0.073  0.004
exper        0.013  0.001
female      -0.454  0.030
married      0.056  0.030

Auxiliary parameter(s):
      Median MAD_SD
sigma 0.485  0.009

Sample avg. posterior predictive distribution of y:
         Median MAD_SD
mean_PPD 1.661  0.020

------
* For help interpreting the printed output see ?print.stanreg
* For info on the priors used see ?prior_summary.stanreg
```

计算 R^2

```
summary(bayes_R2(post))
  Min.    1st Qu.  Median    Mean    3rd Qu.     Max.
  0.2811  0.3219   0.3353    0.3345  0.3463      0.3875
```

Outline

1. 传统多元线性回归模型
2. 贝叶斯多元线性回归模型：理论
3. 贝叶斯多元线性回归模型：WinBUGS
4. 贝叶斯多元线性回归模型：RStan
5. 用 bayesplot 进行可视化
6. 先验分布的设定

bayesplot

The *bayesplot* package provides a variety of ggplot2-based plotting functions for use after fitting Bayesian models (typically, though not exclusively, via Markov chain Monte Carlo).

The plotting functions in *bayesplot* are organized into several modules:

- MCMC: Visualizations of Markov chain Monte Carlo (MCMC) simulations generated by any MCMC algorithm as well as diagnostics. There are also additional functions specifically for use with models fit using the No-U-Turn Sampler (NUTS).
- PPC: Graphical posterior predictive checks (PPCs).

To use the posterior draws we'll extract them from the fitted model object:

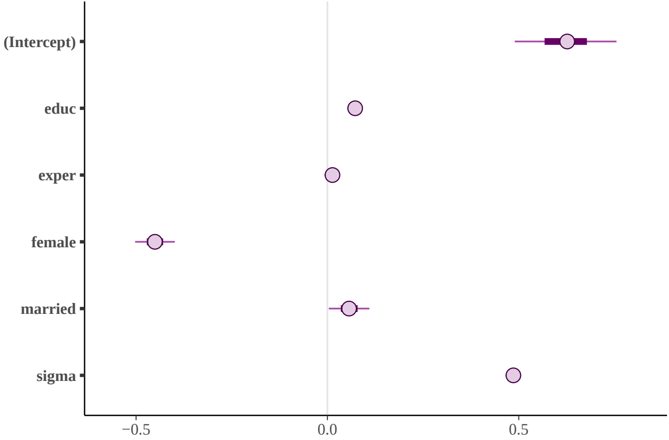
```
library(bayesplot)
post_mc<- as.array(post)
```

Outline

5. 用 bayesplot 进行可视化
   - Rstanarm 的结果展示
   - MCMC 收敛性检查
   - 用 ShinyStan 给出交互式结果

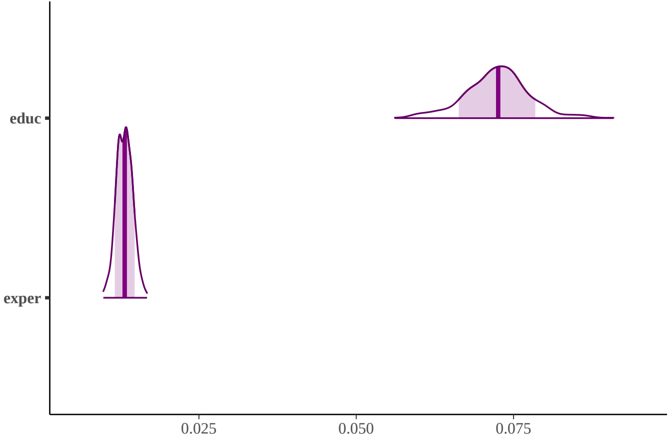
参数的图示：mcmc_intervals

```
color_scheme_set("purple")
mcmc_intervals(post_mc)
```

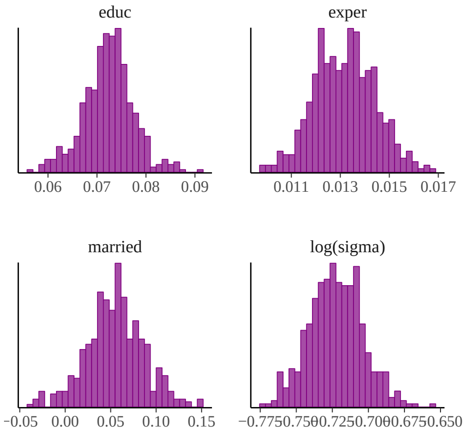


参数的图示：mcmc_areas

```
mcmc_areas(post_mc, prob = 0.8, pars = c("educ","exper"))
```

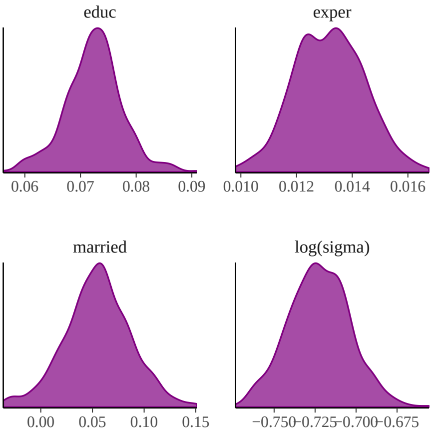


```
mcmc_hist(post_mc,pars=c("educ","exper","female",
 "married","sigma"),transformations=list("sigma"="log"))
```



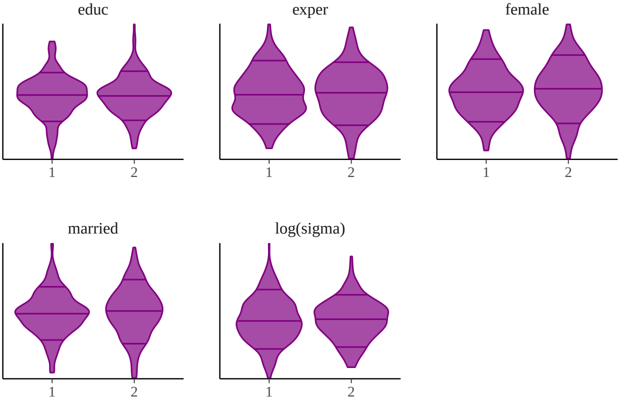
后验参数分布密度图

```
mcmc_dens(post_mc, pars = c("educ","exper","female",
"married","sigma"),transformations=list("sigma"="log"))
```



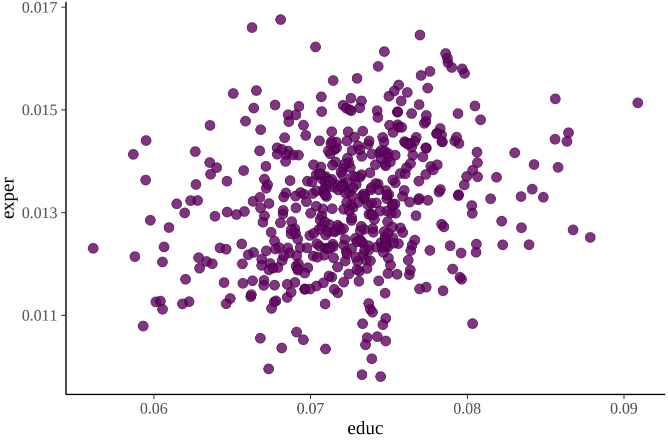
参数估计的小提琴图

```
mcmc_violin(post_mc,pars=c("educ","exper","female",
"married","sigma"),transformations=list("sigma"="log"),
probs=c(0.1,0.5,0.9))
```



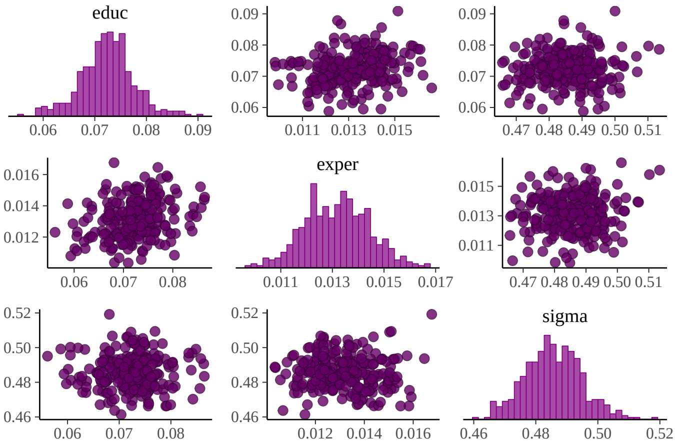
两个参数的相关性：散点图

```
mcmc_scatter(post_mc, pars = c("educ","exper"))
```



多个参数的两两比较

```
mcmc_pairs(post_mc, pars = c("educ","exper","sigma"))
```



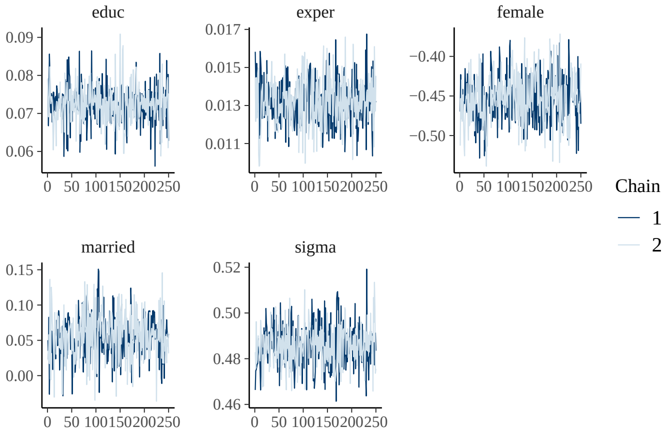
Outline

5 用 bayesplot 进行可视化

- Rstanarm 的结果展示
- MCMC 收敛性检查
- 用 ShinyStan 给出交互式结果

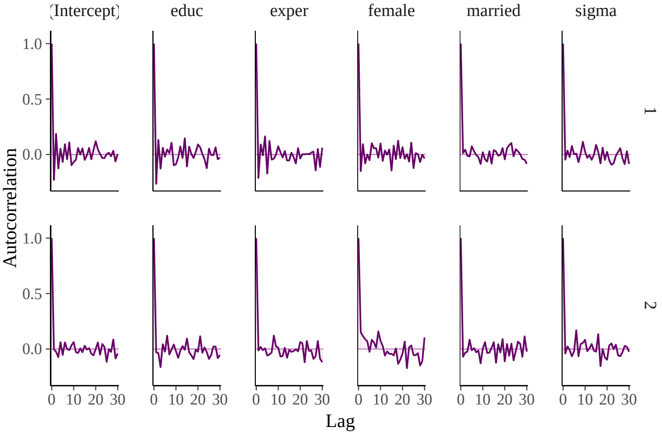
MCMC 收敛性：Trace

```
color_scheme_set("mix-blue-red")
mcmc_trace(post_mc,pars=c("educ","exper","female",
  "married","sigma"), facet_args = list(nrow = 2))
```

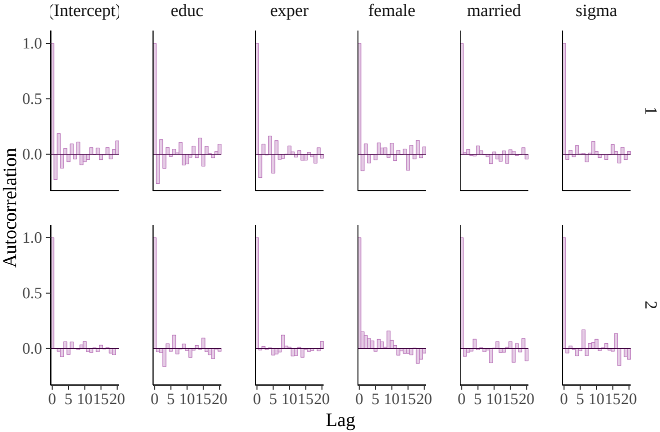


MCMC 收敛性：自相关（acf）

```
mcmc_acf(post_mc, lags = 30)
```

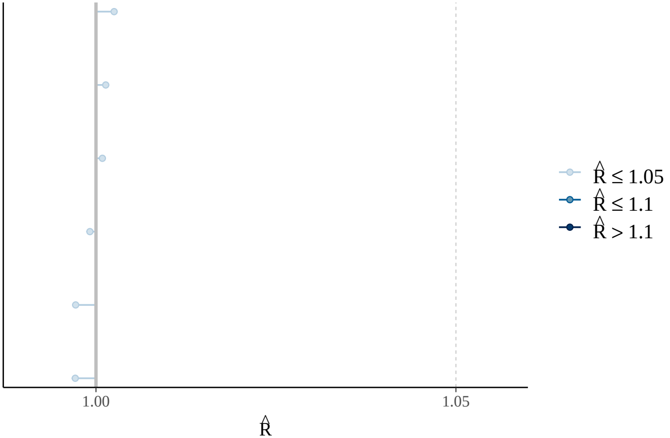


```
mcmc_acf_bar(post_mc)
```



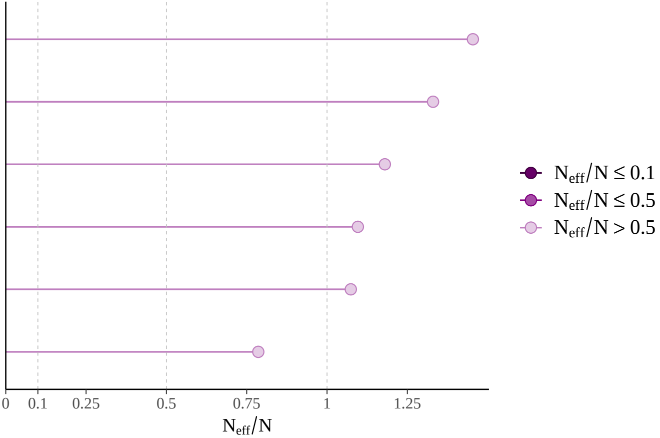
MCMC 收敛性：缩减因子

```
rhats <- rhat(post)
mcmc_rhat(rhats)
```



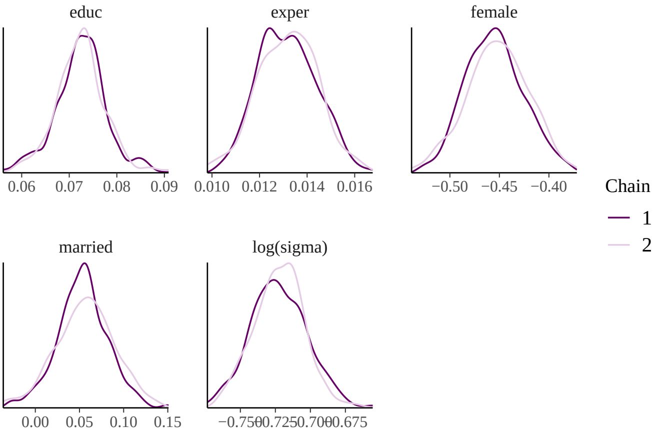
MCMC 收敛性：有效样本量

```
plot(post,"ess",size = 3)
```



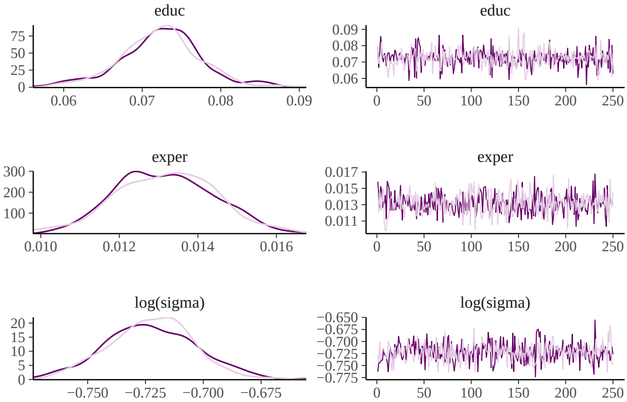
多链条参数后验分布密度曲线

```
mcmc_dens_overlay(post_mc,pars=c("educ","exper","female",
"married","sigma"),transformations=list("sigma"="log"))
```



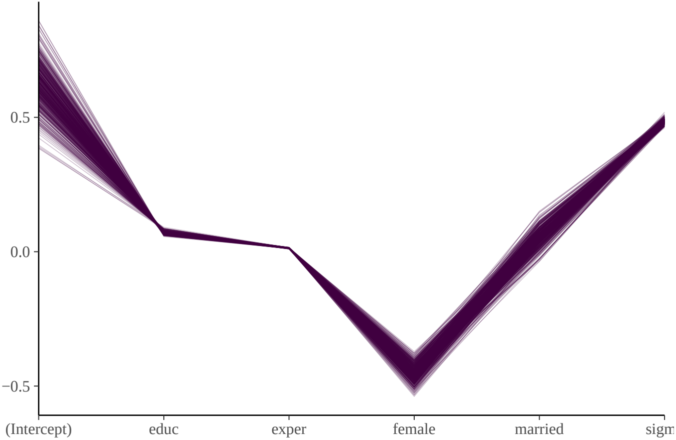
两种图形放在一起

```
mcmc_combo(post_mc,
  combo = c("dens_overlay", "trace"),
  pars = c("educ","exper", "sigma"),
  transformations = list(sigma = "log"),
  gg_theme = legend_none())
```



平行坐标图

```
mcmc_parcoord(post_mc)
```



Outline

5. 用 bayesplot 进行可视化
   - Rstanarm 的结果展示
   - MCMC 收敛性检查
   - 用 ShinyStan 给出交互式结果

launch_shinystan(post)

Shinystan 给出的结果

Save & Close

SHINYSTAN DIAGNOSE ESTIMATE EXPLORE MORE

Parameters plot | Posterior summary statistics | Generate LaTeX table

Digits 3

Copy | Print | Download | Column visibility | Show 10 entries

Glossary

Regex searching

	n_eff	Rhat	mean	mcse	sd	2.5%	25%	50%	75%	97.5%
(Intercept)	4,986	1	0.624	0.001	0.081	0.463	0.568	0.627	0.678	0.781
educ	5,388	1	0.073	0	0.005	0.062	0.069	0.073	0.076	0.083
exper	5,053	1	0.013	0	0.001	0.011	0.012	0.013	0.014	0.016
female	4,782	1	-0.451	0	0.032	-0.513	-0.472	-0.451	-0.429	-0.39
married	4,478	1	0.057	0	0.032	-0.006	0.035	0.057	0.079	0.119
sigma	5,389	1	0.486	0	0.01	0.468	0.48	0.486	0.492	0.506
mean_PPD	5,005	0.999	1.659	0	0.02	1.621	1.646	1.659	1.672	1.699
log-posterior	1,814	1	-886.569	0.042	1.808	-890.922	-887.511	-886.245	-885.246	-884.107

Showing 1 to 8 of 8 entries

First Previous Next Last

Outline

1 传统多元线性回归模型

2 贝叶斯多元线性回归模型：理论

3 贝叶斯多元线性回归模型：WinBUGS

4 贝叶斯多元线性回归模型：RStan

5 用 bayesplot 进行可视化

6 先验分布的设定

Rstanarm 默认的先验分布是弱信息先验

一般情况下，Rstanarm 所指定的默认先验分布不是无信息先验，而是**弱信息先验** (weakly informative).
Rstanarm 参数先验分布的指定：

- **prior_intercept** : Model intercept, after **centering** predictors (Note: the user does not need to manually center the predictors.)
  - prior_intercept = normal(location = 0, scale = 10)
- **prior**: Regression coefficients. Does not include coefficients that vary by group in a multilevel model.
  - prior = normal(location = [0,0], scale = [2.5,2.5])
- **prior_aux**: Auxiliary parameter, e.g. error SD (interpretation depends on the GLM).
  - prior_aux = exponential(1)
- **prior_covariance**: Covariance matrices in multilevel models with varying slopes and intercepts.
- See help('prior_summary.stanreg') for more details

Rstanarm 先验分布的比例调节

- 查看先验：prior_summary(post)
- Automatic prior scale adjustments
- 取消自动调整：autoscale = FALSE
- 均匀（无信息）先验：prior = NULL

```
Priors for model 'post'
------
Intercept (after predictors centered)
 ~ normal(location = 0, scale = 10)
     **adjusted scale = 5.95

Coefficients
 ~ normal(location = [0,0,0,...], scale = [2.5,2.5,2.5,...])
     **adjusted scale = [0.57,0.12,1.49,...]

Auxiliary (sigma)
 ~ exponential(rate = 1)
     **adjusted scale = 0.59 (adjusted rate = 1/adjusted scale)
------
See help('prior_summary.stanreg') for more details
```

再计算：Rstanarm 无信息先验

模型：

$$\log(wage)_i \sim N(\mu_i, \sigma^2)$$
$$\mu_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 female_i + \beta_4 married_i$$

无信息先验分布：

$\beta_i \sim dflat(), i = 0, 1, 2, 3, 4$
$\sigma^2 \sim IG(10^3, 10^3)$
在 Rstanarm 中，如下代码指定均匀先验：

```
post_pr <- stan_glm(
  lwage ~ educ + exper + female + married,
  data = wage_data,
  family = gaussian(link = "identity"),
  prior = NULL,
  prior_intercept = NULL,
  prior_aux = NULL
)
```

再计算：Rstanarm 指定有信息先验

```
beta_prior <- normal(
  location = c(0, 0, -1, 0),
  scale = c(1, 1, 1, 1),
  autoscale = FALSE
)
post_prior <- stan_glm(
  lwage ~ educ + exper + female + married,
  data = wage_data,
  family = gaussian(link = "identity"),
  prior = beta_prior,
  prior_intercept = normal(0,1),
  prior_aux = cauchy(0,3)
)
```

查看 Rstanarm 的先验分布

prior_summary(post_prior)

结果:

```
Priors for model 'post_prior'
------
Intercept (after predictors centered)
 ~ normal(location = 0, scale = 1)
     **adjusted scale = 0.59
Coefficients
 ~ normal(location = [ 0, 0,-1,...], scale = [1,1,1,...])
Auxiliary (sigma)
 ~ half-cauchy(location = 0, scale = 3)
     **adjusted scale = 1.78
------
See help('prior_summary.stanreg') for more details
```

总结

1. 传统多元线性回归模型
   - 模型假设
   - 参数估计
   - 模型诊断
2. 贝叶斯多元线性回归模型
   - 无信息先验
   - 共轭先验
   - G-先验
3. 运用 WinBUGS
4. 运用 RStan
   - 用 Rstanarm 建立模型、抽取 MCMC 样本、给出结果
   - Rstanarm 的先验分布
   - 用 Bayesplot 进行可视化
   - 用 ShinyStan 给出交互式结果展示