治理技术专题

定量政治分析方法

Quantitative Analysis II

苏 毓 淞

清华大学社会科学院政治学系

第十四讲　多层次回归

MLM 发展背景

- 1980 年代中后期，统计学家关注到层次结构数据 (或称嵌套数据，nested data)，并提出一系列嵌套数据的统计分析方法，统称为“多层次模型”(multilevel model)。
- 多层次模型 (multilevel model) 最先应用于教育学领域，后用于心理学、社会学、政治学、经济学、组织行为与管理科学等领域，逐步应用到医学及公共卫生等领域。

MLM 不同的名称

- 分层线性模型（Hierarchical Linear Model）
- 多层次（水平）分析（Multilevel Analysis）
- 多层次回归模型（Multilevel regression models）
- 混合模型（Mixed Models）
- 随机系数模型（Random Coefficient Models）
- 方差成分模型（Variance Component Model）

Varying Intercept and Varying Slope models

- LSDV 模型：回归中加入集体层次 (group level) 的虚拟变量，这些虚拟变量的斜率可以当成截距解读，所以这模型又叫做 Varying Intercept 模型。

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i$$

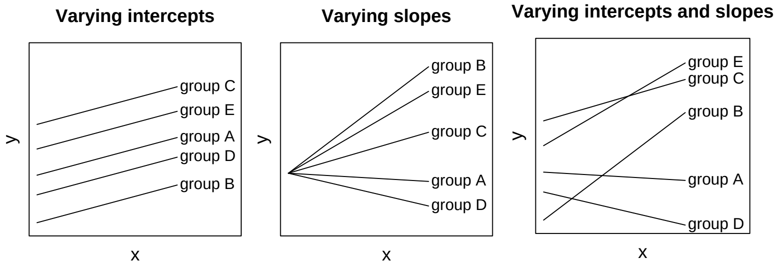
- 如果只让 β 也随集体层次变动，则称为 Varying slope 模型。

$$y_i = \alpha + \beta_{j[i]} x_i + \epsilon_i$$

- 如果只 α, β 都随集体层次变动，则称为 Varying-intercept Varying-slope 模型。

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i$$

Varying Intercept and Varying Slope models



数据类型

- 面板数据 Panel data
- TSCS 数据
- 非嵌套数据 Non-nested data

模型选择

- LSDV
- 固定效应模型和随机效应模型
- 多层次回归 Multilevel regressions.

固定效应模型和随机效应模型

- Varying-intercept Varying-slope 模型有时候也被称之为随机效应模型 (或者随机系数模型)，因为集体层次的截距和斜率是从概率分布中的随机项。
- 固定效应指的是未建模的 Varying-intercepts(没有自变量估计它)，但是有时候有指的是截距不变的回归模型。
- 多层次回归模型综合了两个模型的优点，所以请忘了固定效应模型和随机效应模型！

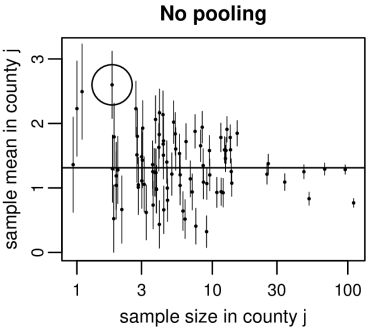
多层次回归模型

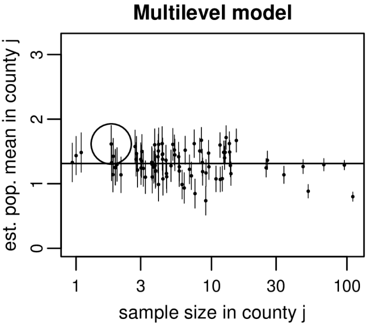
- 如果集体层次的变异量为零，则多层次回归模型等于一般回归模型。
- 如果集体层次量太少，则多层次回归模型和 LSDV 模型差异不大。
- 不过，这也说明了，总是使用多层次回归模型不会错！

池化 pooling

- 如果针对集体层次量 J，进行 J 个回归模型，称之为 no pooling（无池化）分析。
- 如果针对集体层次量 J，进行 1 个回归模型，称之为 ncomplete pooling（全池化）分析。
- 多层次回归是以上两种的综合（加权平均），称之为部分池化分析 (partial pooling)

池化图示


No pooling
sample mean in county j
0
1
2
3
1
3
10
30
100
sample size in county j


Multilevel model
est. pop. mean in county j
0
1
2
3
1
3
10
30
100
sample size in county j

部分池化的估计值

$$y_i = \alpha_{j[i]} + \epsilon_i$$

- 多层次回归的估计值是以上两种的综合（加权平均）。

$$\hat{\alpha}_j^{\text{多层次模型}} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{\text{all}}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}},$$

- 如果某个集体层次的样本量小，则 $\hat{\alpha}_j$ 会被拉往全池化估计值 $\bar{y}_{\text{all}}$。某个集体层次的样本量为零，则 $\hat{\alpha}_j = \bar{y}_{\text{all}}$
- 某个集体层次的样本量大，则 $\hat{\alpha}_j$ 会被拉往无池化估计值 $\bar{y}_j$。某个集体层次的样本量 $\to \infty$，则 $\hat{\alpha}_j = \bar{y}_j$

无池化和全池化模型问题

- 全池化模型忽略了集体层次的变异。
- 无池化模型则容易因样本量少时出现极值。
- 部分池化（多层次回归）模型则考虑到了集体层次的变异，同时，将因样本量少产生的极值，牵引到全池化的估计值上。

使用 R 应用多层次回归

- varying intercepts model without predictors:

```
R> M0 <- lmer(y ~ 1 + (1|group))
R> display(M0)
lmer(formula = y ~ 1 + (1 | group))
coef.est  coef.se
   -0.49     0.84

Error terms:
 Groups   Name        Std.Dev.
 group    (Intercept) 2.32
 Residual             4.09
---
number of obs: 100, groups: group, 10
AIC = 583.6, DIC = 580.5
deviance = 579.0
```

多层次回归数学表达式

- varying intercepts model
- 个体层次：(两种写法)
  - $y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_y^2)$
  - $y_i \sim N(\alpha_j + \beta x_i, \sigma_y^2)$
- 集体层次 (无自变量)：(两种写法)
  - $\alpha_j \sim N(\mu_a, \sigma_\alpha^2)$
  - $\alpha_j = \mu_\alpha + \eta_j, \qquad \eta_j \sim N(0, \sigma_\alpha^2)$
- 集体层次 (有自变量)：(两种写法)
  - $\alpha_j \sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2)$
  - $\alpha_j = \gamma_0 + \gamma_1 u_j + \eta_j, \qquad \eta_j \sim N(0, \sigma_\alpha^2)$

多层次回归数学表达式

- Varying intercepts and varying slopes model without group level predictors

$$y_i \sim \mathcal{N}(\alpha_{j[i]} + \beta_{j[i]} x_i,\ \sigma_y^2),\ \text{for } i = 1, \ldots, n$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right),\ \text{for } j = 1, \ldots, J,$$

- Varying intercepts and varying slopes model with group level predictors

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \gamma_0^\alpha + \gamma_1^\alpha u_j \\ \gamma_0^\beta + \gamma_1^\beta u_j \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right),\ \text{for } j = 1, \ldots, J.$$

使用 R 应用多层次回归

- varying intercepts model with one predictor:

```
R> M1 <- lmer(y ~ x + (1|group))
R> display(M1)
lmer(formula = y ~ x + (1 | group))
            coef.est  coef.se
(Intercept) -0.71     0.83
x            2.08     0.36

Error terms:
 Groups   Name        Std.Dev.
 group    (Intercept) 2.38
 Residual             3.51
---
number of obs: 100, groups: group, 10
AIC = 557.1, DIC = 551.5
deviance = 550.3
```

使用 R 应用多层次回归

- varying intercepts model with one predictor:

```
R> M2 <- lmer(y ~ x + (1+x|group))
R> display(M2)
lmer(formula = y ~ x + (1 + x | group))
            coef.est coef.se
(Intercept) -0.75     0.80
x            2.45     0.58

Error terms:
 Groups   Name        Std.Dev. Corr
 group    (Intercept) 2.30
          x           1.48     0.75
 Residual             3.18
---
number of obs: 100, groups: group, 10
AIC = 549.5, DIC = 540.9
deviance = 539.2
```

使用 R 应用多层次回归

- varying intercepts model with group level predictor:

```
> M3 <- lmer(y ~ x + u.all + (1+x|group))
> display(M3)
lmer(formula = y ~ x + u.all + (1 + x | group))
            coef.est coef.se
(Intercept)  1.55     0.56
x            2.67     0.46
u.all       -0.34     0.22

Error terms:
 Groups   Name        Std.Dev. Corr
 group    (Intercept) 0.08
          x           1.11     1.00
 Residual             2.75
---
number of obs: 100, groups: group, 10
AIC = 507.9, DIC = 490.7
deviance = 492.3
```

讨论

- J 太少时，多层次回归模型的优势发挥不出来，无法良好的估计组间方差（J 太小），但至少不会比 no pooling 的回归差。
- $J=2$ 时，跟传统加上虚拟变量的回归没有两样。
- 即便是每组只有 1 个样本量，也足以使用多层次回归。

Bayeisan Multilevel Modeling

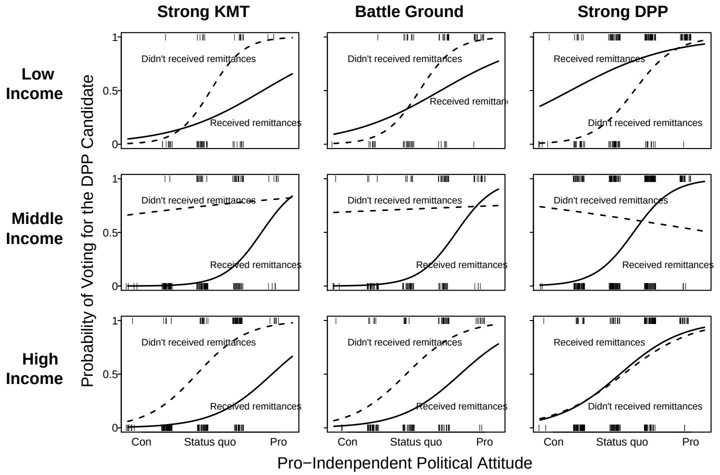
$$y_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]}, \sigma_y^2\right)$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right)$$

- non-informative priors

$$\begin{aligned} \mu_\alpha &\sim N(0, 1000) \\ \mu_\beta &\sim N(0, 1000) \\ \sigma_y &\sim U(0, 1000) \\ \sigma_\alpha &\sim U(0, 1000) \\ \sigma_\beta &\sim U(0, 1000) \\ \rho &\sim U(-1, 1) \end{aligned}$$

案例 1: Multilevel Logistic Regression

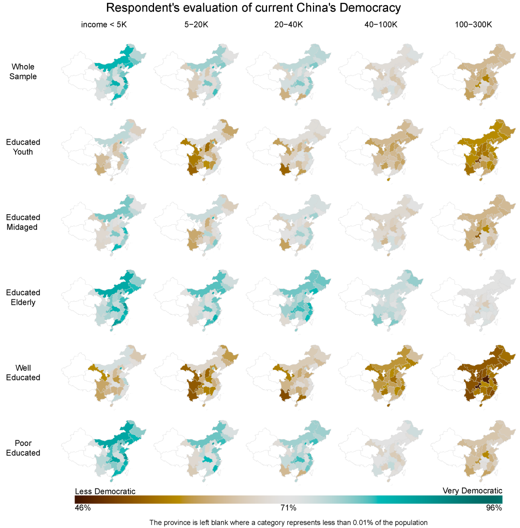


案例 1: Multilevel Logistic Regression

```
fit <- glmer(y ~ 1 + x1 + x2 + x1:x2 +
(1 + x1 + x2 + x1:x2|income) +
(1 + x1 + x2 + x1:x2|region), family=binomial(link="logit"), data=dat)
```

- y: binary variable
- x1: continuous variable
- x2: binary variable
- income: 3 levels categorical variable
- region: 3 levels categorical variable

案例 2: MRP


Respondent's evaluation of current China's Democracy
income < 5K
5−20K
20−40K
40−100K
100−300K
Whole Sample
Educated Youth
Educated Midaged
Educated Elderly
Well Educated
Poor Educated
Less Democratic
Very Democratic
46%
71%
96%
The province is left blank where a category represents less than 0.01% of the population

案例 2: MRP

```
fit <- glmer (y ~ gender*gdp +
  (1 + gender | region) +
  (1 + gender | province) +
  (1 | gender.region) +
  (1 | gender.province) +
  (1 | gender) +
  family=binomial(link="logit"))
```