

## 相加混合模型

本模块使用广义混合模型或相加混合模型 (GAMM: generalize additive mixed model) 检验危险因素 (X) 和结果变量 (Y) 之间的联系。相加混合模型是广义相加模型与混合模型的结合, 在广义相加模型中可以指定平滑曲线拟合项, 在混合模型中可以引进随机效应 (可以是随机截距或/和随机时间斜率)。它主要是用于重复测量资料的分析。数据中一般有时间变量 (T), 结果变量随时间 (T) 变化, 同时所研究的危险因素 (X) 又对结果变量 (Y) 可能有影响作用。危险因素 (X) 可以是固定的也可以随时间 (T) 变化。如果危险因素 (X) 随时间 (T) 变化, 每次重复测量都测量了危险因素 (X) 与结果变量 (Y) 的值, 还可以分析危险因素 (X) 的滞后效应, 反映危险因素 (X) 变化在前, 结果 (Y) 变化在后。滞后效应分析要结合数据操作菜单下在多条记录内计算统计量模块, 首先对数据进行预处理, 然后用该模块分析。建立回归模型的目的是把危险因素 (X) 的作用从时间 (T) 的作用中区别出来, 分析危险因素 (X) 是否有作用? 是什么样的作用?

### 如何模拟时间变量 (T) 与结果变量 (Y) 的关系:

可以用二种方式拟合时间 (T) 与结果变量 (Y) 的关系:

1. 不假定时间变量 (T) 与结果变量 (Y) 是直线的关系或是二次项的曲线 (时间加上时间平方项) 的关系, 用平滑曲线拟合时间 (T) 与结果变量 (Y) 的关系。此时时间变量 (T) 不能被指定为随机效应, 模型的随即效应仅限于随机截距。
2. 如果结果变量 (Y) 随时间呈直线性变化, 用时间变量 (T) 的一次项来拟合。如果时间变量 (T) 与结果变量 (Y) 的关系不完全是直线关系, 在模型中要考虑是否应引进时间的平方项 ( $T^2$ )。如果每个个体的重复测量时间是固定的, 时间变量 (T) 是个分类型的变量, 也可以按分类型变量拟合时间 (T) 的作用。

### 如何拟合危险因素 (X) 与结果变量 (Y) 的关系:

1. 如果危险因素 (X) 是固定的 (一般是基线测量值), 主要看危险因素 (X) 与时间 (T) 的交互作用, 如果有交互作用表示危险因素 (X) 不同, 结果变量 (Y) 随时间 (T) 的变化速度 (斜率) 不同。
2. 如果所研究的危险因素 (X) 随时间 (T) 变化, 这时时间 (T) 与结果 (Y) 有关, 又与危险因素 (X) 有关, 模型中需要调整时间 (T) 的作用才能正确观察危险因素 (X) 的作用。因此对时间 (T) 的作用拟合得充分与否直接影响对危险因素 (X) 作用的评估是否正确, 参考上面如何拟合时间 (T) 的作用。拟合危险因素 (X) 的作用可以首先用平滑曲线 (使用基本统计菜单下的广义相加混合模型曲线拟合 X), 观察危险因素 (X) 与结果 (Y) 是否是直线性关系, 然后考虑是否可以采用 (1) 直线拟合; (2) 加上二次项的曲线拟合; (3) 把危险因素 (X) 分组, 按分类型变量拟合。

如果暴露 (X) 不随时间变化, 如何拟合暴露 (X) 与时间 (T), 系统给出如下模型供选择, 并可自动比较几个常用模型。

1) 如果 T 是连续性变量

(1)  $Y = X + T$

(2)  $Y = X + T + T^2$

(3)  $Y = X + T + X*T$

(4)  $Y = X + T + T^2 + X*T + X*T^2$

(5)  $Y = X + s(T)$

2) 如果 X 是分类变量

(1)  $Y = X + \text{factor}(T)$

(2)  $Y = X + \text{factor}(T) + X*\text{factor}(T)$

其中: Y= 结果变量, X=危险因素, T=时间,  $T^2$ =时间的平方项, C=其它协变量

系统自动用[似然比检验](#)进行模型之间的比较, 判断如何拟合时间 (T) 及是否存在 X 与 T 的交互作用。

例 1 (输入界面) 对两组小白鼠分别与 0、1、2、4 周测量了某指标 (固定时点测量的数据) 原始数据如下, 分析组间差异:

| Group | 0    | 1    | 2    | 4    |
|-------|------|------|------|------|
| 0     | 0.35 | 1.01 | 1.47 | 2.46 |
| 0     | 0.77 | 1.32 | 1.6  | 2.54 |
| 0     | 0.48 | 1.18 | 1.65 | 2.86 |
| 0     | 0.63 | 1.42 | 1.88 | 3.13 |
| 1     | 0.45 | 0.59 | 0.64 | 0.99 |
| 1     | 0.56 | 0.86 | 1.37 | 2.04 |
| 1     | 1.08 | 1.44 | 1.93 | 2.63 |
| 1     | 0.55 | 1.2  | 1.68 | 2.87 |

首先将数据转换成纵向结构如下:

| GROUP | ID | VT | Y    |
|-------|----|----|------|
| 0     | 1  | 0  | 0.35 |
| 0     | 1  | 1  | 1.01 |
| 0     | 1  | 2  | 1.47 |
| 0     | 1  | 4  | 2.46 |
| 0     | 2  | 0  | 0.77 |

|     |     |     |      |
|-----|-----|-----|------|
| ... | ... | ... | ...  |
| 1   | 8   | 4   | 2.87 |

然后调用该模块，输入界面如下：

### 重复测量数据GAMM分析 ?

**标题：**

**选择分析对象：**

**应变变量(Y)**

| 变量名   | 分布类型     | 联系函数     |
|-------|----------|----------|
| Value | Gaussian | Identity |

**暴露变量(X)**

| 变量    | 研究对象编号ID |
|-------|----------|
| GROUP | id       |

**时间变量 T**

**选择随机效应**

**模型构建(M)**

**拟合暴露与时间**

**调整变量**  含未调整模型

| 变量 | 曲线拟合 |
|----|------|
|    |      |

**选择分层变量**

**输出内容与格式**

**精确到小数点：**

刷新
保存
查看结果

输出结果：

结果变量: Value  
 暴露变量: GROUP  
 调整变量: None

Exposure: GROUP

|                         |                         |
|-------------------------|-------------------------|
| Outcome: Value          |                         |
| factor(GROUP)1          | 0.1 (-0.5, 0.6) 0.848   |
| tmp.TIME                | 0.5 (0.5, 0.6) <0.001   |
| factor(GROUP)1:tmp.TIME | -0.2 (-0.3, -0.1) 0.006 |

Results in table:  $\beta$  (95%CI) Pvalue / OR (95%CI) Pvalue

此表用易侓统计软件 (www.empowerstats.com) 和 R 软件生成, 生成日期: 2016-04-18

各模型所用的样本量

| Exposure | Outcome | Model | N obs | N Group |
|----------|---------|-------|-------|---------|
| GROUP    | Value   | Crude | 32    | 8       |

Detailed outputs:

Outcome: Value

Exposure: GROUP

|                                  | Model 1                    | Model 2                   | LRT p-value |
|----------------------------------|----------------------------|---------------------------|-------------|
| (Intercept)                      | 0.6 (0.1, 1.0)<br>0.017    | 0.7 (0.4, 1.1)<br>0.001   |             |
| factor(GROUP)1                   | 0.1 (-0.5, 0.7)<br>0.743   | -0.2 (-0.7, 0.2)<br>0.360 |             |
| factor(tmp.TIME)1                | 0.7 (0.3, 1.0)<br>0.001    | 0.5 (0.2, 0.8)<br>0.001   |             |
| factor(tmp.TIME)2                | 1.1 (0.8, 1.4)<br><0.001   | 0.9 (0.6, 1.2)<br><0.001  |             |
| factor(tmp.TIME)4                | 2.2 (1.9, 2.5)<br><0.001   | 1.8 (1.6, 2.1)<br><0.001  |             |
| factor(GROUP)1:factor(tmp.TIME)1 | -0.3 (-0.8, 0.2)<br>0.214  | NA                        |             |
| factor(GROUP)1:factor(tmp.TIME)2 | -0.3 (-0.8, 0.1)<br>0.169  | NA                        |             |
| factor(GROUP)1:factor(tmp.TIME)4 | -0.7 (-1.2, -0.2)<br>0.008 | NA                        |             |
| ---                              |                            |                           |             |
| Log likelihood                   | -4.313                     | -9.087                    | 0.023       |

|                | Model 3               | Model 4                | LRT p-value |
|----------------|-----------------------|------------------------|-------------|
| (Intercept)    | 0.6 (0.2, 1.0) 0.009  | 0.7 (0.4, 1.1) 0.001   |             |
| factor(GROUP)1 | 0.1 (-0.5, 0.6) 0.809 | -0.2 (-0.7, 0.2) 0.352 |             |
| tmp.TIME       | 0.6 (0.3, 0.9) 0.001  | 0.5 (0.2, 0.7) 0.001   |             |

|                          |                           |                       |       |
|--------------------------|---------------------------|-----------------------|-------|
| tmp.TIME2                | 0.0 (-0.1, 0.1) 0.743     | 0.0 (-0.1, 0.0) 0.826 |       |
| factor(GROUP)1:tmp.TIME  | -0.2 (-0.6, 0.2)<br>0.318 | NA                    |       |
| factor(GROUP)1:tmp.TIME2 | 0.0 (-0.1, 0.1) 0.834     | NA                    |       |
| ---                      |                           |                       |       |
| Log likelihood           | -4.750                    | -9.208                | 0.012 |

|                         |                         |                        |             |
|-------------------------|-------------------------|------------------------|-------------|
|                         | Model 5                 | Model 6                | LRT p-value |
| (Intercept)             | 0.6 (0.3, 1.0) 0.003    | 0.8 (0.4, 1.1) <0.001  |             |
| factor(GROUP)1          | 0.1 (-0.5, 0.6) 0.848   | -0.2 (-0.7, 0.2) 0.344 |             |
| tmp.TIME                | 0.5 (0.5, 0.6) <0.001   | 0.5 (0.4, 0.5) <0.001  |             |
| factor(GROUP)1:tmp.TIME | -0.2 (-0.3, -0.1) 0.006 | NA                     |             |
| ---                     |                         |                        |             |
| Log likelihood          | -4.819                  | -9.237                 | 0.003       |

| Model | LLK    | LRT   | LRT   | Model | LLK    | LRT   | LRT   | LRT   |
|-------|--------|-------|-------|-------|--------|-------|-------|-------|
| 1     | -4.313 | Ref.  |       | 2     | -9.087 | Ref.  |       | 0.023 |
| 3     | -4.750 | 0.646 | Ref.  | 4     | -9.208 | 0.623 | Ref.  | 0.012 |
| 5     | -4.819 | 0.908 | 0.934 | 6     | -9.237 | 0.861 | 0.812 | 0.003 |

### 结果解释

首先给出最终被选模型： $Y = \text{GROUP} + \text{tmp.TIME} + \text{GROUP} * \text{tmp.TIME}$

tmp.TIME 表示时间变量，该模型中时间按连续性变量分析，tmp.TIME 的回归系数 0.5 表示 GROUP=0 组，Y 随 T 的变化为每周增加 0.5，95%可信区间 0.5-0.6,  $P < 0.001$ 。GROUP\*tmp.TIME 项为交互作用项，回归系数 -0.2 表示 GROUP=1 组 Y 随时间变化的斜率与 GROUP=0 组的差为 0.2， $P = 0.005$ ，即每周变化为  $0.5 - 0.2 = 0.3$ 。这个斜率差值反映两组的差异。

下面给出模型间的比较的似然比检验过程。

### 结果解释

以模型 1 为例，模型 1 是将时间 (T) 按分类变量分析：

intercept 项回归系数 0.6 表示 GROUP=0，T=0 点 Y 的测量值的均数，其  $P = 0.017$  表示与 0 比差异显著；

GROUP 项的回归系数 0.1 表示 T=0 时点，GROUP=1 与 GROUP=0 两组 Y 的差异为 0.1， $p = 0.743$ ，差异不显著；

factor (T) 1 项回归系数 0.7 表示 GROUP=0 组、T=1 与 T=0 两个时点的差异为 0.7， $p < 0.001$ ；

factor (T) 2 项回归系数 1.1 表示 GROUP=0 组、T=2 与 T=0 两个时点的差异为 1.1， $p < 0.001$ ；

factor (T) 4 项回归系数 2.2 表示 GROUP=0 组、T=4 与 T=0 两个时点的差异为 2.2， $p < 0.001$ ；

GROUP\*factor (T) 1 项的回归系数-0.3 表示 T=1 与 T=0 两个时点的差，GROUP=1 组比 GROUP=0 组少 0.3， $p=0.214$ ；

GROUP\*factor (T) 2 项的回归系数-0.3 表示 T=2 与 T=0 两个时点的差，GROUP=1 组比 GROUP=0 组少 0.3， $p=0.169$ ；

GROUP\*factor (T) 4 项的回归系数-0.7 表示 T=4 与 T=0 两个时点的差，GROUP=1 组比 GROUP=0 组少 0.7， $p=0.008$ ；

上面三个是交互作用项，反映两组时点变化的差异，即时间与分组的交互作用。

模型 2 与模型 1 相比，剔除了 GROUP 与 factor (T) 的交互作用项。结果似然比检验  $p=0.023$  有显著差异，表示不能用模型 2 替代模型 1。

模型 3 把时间 T 按连续性变量，用 T 加  $T^2$  拟合 Y 与 T 的关系，并有 X 与 T、X 与  $T^2$  的交互作用项。模型 4 从模型 3 中剔除交互作用项。两模型比较的似然比检验  $p=0.012$ ，差异显著，表示不能用模型 4 替代模型 3。

模型 5 是从模型 3 中剔除  $T^2$  项，直线拟合 T。模型 6 是在模型 5 的基础上剔除 GROUP 与 T 的交互作用项。两模型比较的似然比检验  $p=0.0003$ ，差异显著，表示不能用模型 6 替代模型 5。

最后一个表是比较模型 1、3、5，模型 3 与模型 1 相比，LRT（似然比检验） $p=0.646$ ，差异不显著，模型 5 与模型 1 相比，似然比检验  $p=0.908$ ，差异不显著。模型 5 与模型 3 相比， $p=0.934$  差异不显著。这些比较的结果表示模型 5 最佳。同理系统自动对模型 2、4、6 之间也作了相互比较。基于前面的结果，无交互作用模型不能替代有交互作用的模型，因此这部分结果可以忽略。

例 2. 分析几个国家逐年 malaria incidence 与男、女肿瘤死亡率 (cancermortality.male, cancernortality.female) 的关系 (危险因素 X 与结果变量 Y 均随时间变化)，输入界面如下：

**重复测量数据GAMM分析** ?

标题:

选择分析对象:

应变量(Y)

| 变量名                    | 分布类型     | 联系函数     |
|------------------------|----------|----------|
| CANCERMORTALITY.MALE   | Gaussian | Identity |
| CANCERMORTALITY.FEMALE | Gaussian | Identity |

暴露变量(X)

| 变量                |
|-------------------|
| MALARIA INCIDENCE |

研究对象编号ID:

时间变量 T:

选择随机效应:

模型构建(M):

拟合暴露与时间:

调整变量  含未调整模型

| 变量 | 曲线拟合 |
|----|------|
|    |      |

选择分层变量:

输出内容与格式:

精确到小数点:

输出结果:

结果变量: CANCERMORTALITY.MALE; CANCERMORTALITY.FEMALE

暴露变量: MALARIA INCIDENCE

调整变量: None

Exposure: MALARIA INCIDENCE

|                   | CANCERMORTALITY.MALE           | CANCERMORTALITY.FEMALE         |
|-------------------|--------------------------------|--------------------------------|
| MALARIA INCIDENCE | -0.66 (-0.93, -0.39)<br><0.001 | -0.97 (-1.30, -0.63)<br><0.001 |

Results in table:  $\beta$  (95%CI) Pvalue / OR (95%CI) Pvalue  
 此表用易侖统计软件 (www.empowerstats.com) 和 R 软件生成, 生成日期:  
 2016-06-05

各模型所用的样本量

| Exposure          | Outcome                 | Model | # obs. | # COUNTRY. NEW |
|-------------------|-------------------------|-------|--------|----------------|
| MALARIA INCIDENCE | CANCERMORTALITY. MALE   | Crude | 155    | 5              |
| MALARIA INCIDENCE | CANCERMORTALITY. FEMALE | Crude | 155    | 5              |

Detailed outputs:

Outcome: CANCERMORTALITY. MALE

Exposure: MALARIA INCIDENCE

Sub-group: Total

|                   |                             |
|-------------------|-----------------------------|
| (Intercept)       | 73.09 (41.1, 105.09) <0.001 |
| MALARIA_INCIDENCE | -0.66 (-0.93, -0.39) <0.001 |
| s(tmp.TIME)Fx1    | 2.68 (-4.52, 9.88) 0.466    |

Outcome: CANCERMORTALITY. FEMALE

Exposure: MALARIA INCIDENCE

Sub-group: Total

|                   |                             |
|-------------------|-----------------------------|
| (Intercept)       | 64.94 (55.78, 74.1) <0.001  |
| MALARIA_INCIDENCE | -0.97 (-1.30, -0.63) <0.001 |
| s(tmp.TIME)Fx1    | -1.33 (-7.38, 4.71) 0.666   |

结果解释

本例用平滑曲线拟合时间 (YEAR) 与结果变量 (CANCERMORTALITY. MALE, CANCERMORTALITY. FEMALE) 的关系, 这样充分调整了时间的作用后, 观察危险因素 (MALARIA INCIDENCE) 对结果变量的作用。MALARIA INCIDENCE 每增加 1 个单位, CANCERMORTALITY. MALE 增加-0.66, 95%可信区间-0.93 至-0.39,  $p < 0.001$ ; CANCERMORTALITY. FEMALE 增加-0.97, 95%可信区间-1.30 至-0.63,  $p < 0.001$ 。

曲线拟合 YEAR 与结果变量的关系如下图所示:



