

重复测量数据分析 1.2: X 是连续性的变量

下面看另外一个例子:

例 1.2 给药浓度不同的情况下, 在不同时点 (12-72 小时) 观察的细胞活力数据:

| CONCE | 12 | 24 | 36 | 48 | 72 |
|-------|-------|-------|-------|-------|-------|
| 0 | 99.52 | 99.21 | 98.09 | 99.36 | 98.49 |
| 0 | 99.62 | 99.42 | 99.31 | 99.12 | 99.24 |
| 0 | 98.84 | 98.76 | 99.24 | 99.31 | 99.15 |
| 10 | 96.68 | 92.89 | 86.54 | 76.49 | 65.59 |
| 10 | 95.24 | 91.12 | 85.23 | 77.28 | 65.88 |
| 10 | 95.81 | 90.64 | 84.88 | 76.62 | 64.24 |
| 20 | 89.39 | 76.89 | 69.16 | 64.66 | 51.26 |
| 20 | 89.51 | 75.12 | 68.22 | 62.12 | 50.24 |
| 20 | 88.46 | 75.24 | 68.67 | 62.78 | 50.12 |
| 40 | 76.66 | 67.01 | 55.35 | 50.03 | 44.91 |
| 40 | 73.88 | 68.21 | 55.02 | 48.12 | 42.68 |
| 40 | 75.29 | 66.83 | 56.12 | 48.68 | 44.12 |
| 60 | 68.43 | 56.29 | 46.22 | 38.37 | 33.24 |
| 60 | 66.78 | 54.84 | 44.88 | 37.22 | 32.59 |
| 60 | 68.24 | 55.34 | 45.68 | 36.34 | 33.12 |
| 80 | 59.07 | 41.76 | 34.62 | 26.38 | 10.81 |
| 80 | 58.12 | 39.82 | 35.26 | 24.88 | 10.21 |
| 80 | 58.46 | 41.08 | 34.12 | 24.34 | 11.24 |
| 100 | 42.5 | 27.01 | 18.48 | 7.53 | 3.32 |
| 100 | 43.34 | 28.34 | 17.72 | 8.12 | 3.39 |
| 100 | 42.11 | 28.12 | 17.95 | 8.03 | 3.18 |

这个例子与前例 (1.1) 的区别是, X 不是分类型的观察组, 而是可以看成连续型的给药浓度。如果 X 是一个连续性变量, 它对 Y 的时间变化斜率的影响又是连续性的, 多设置几个 X 的观察点 (如这里设了 7 个浓度), 可以大大提高研究效率。但如果数据分析时没有把 X 当成连续性的, 或没有拟合好 X 与 Y 的关系 (如非直线性的), 则有可能适得其反。

分析模型:

混合线性模型原理参考例 1.1

如果 t 与 Y 的关系是直线性的, X 与 Y 的关系也是直线性的, 方程则非常简单: $Y = \beta_{0i} + \beta_1 * t + \beta_2 * X + \beta_3 * t * X + e$

当然 t 与 Y 不一定是直线性的, 我们可以加入 t^2 项于模型中, 甚至可以加入 t^3 项。方程变为:

$$Y = \beta_{0i} + \beta_1 * t + \beta_2 * t^2 + \beta_3 * X + \beta_4 * t * X + \beta_5 * t^2 * X + e$$

如果不能用 t 、拟合 Y 与 t 的关系, 则只好把 t 当成分类变量了。

同理, X 与 Y 也不一定是直线性的, 我们可以也加入 X^2 (甚至 X^3) 项

$$Y = \beta_{0i} + \beta_1 * t + \beta_2 * X + \beta_3 * X^2 + \beta_4 * t * X + \beta_5 * t * X^2 + e$$

如用到 X^2 则参数解释与应用就复杂多了, 所以通常都尽可能不用。

软件操作

从实验室出来的数据通常可以在 Excel 里整理成上表的格式, 这里第一列是给药浓度, 第一行 CONC 后面是时间值。中间如有数据缺失, 最好用 NA 填上。打开“易侖统计”, 点击“工具”菜单, 选“统计工具箱”, 再点击“统计工具”选“重复测量数据随机效应模型分析”, 把上表数据从 Excel 里复制粘贴进数据输入框内, 如下图所示:

选“第一列为分组因素”再选“将分组因素按连续性变量分析”。

重复测量数据随机效应模型分析

分析项目标题: 输出文件名:

复制粘贴m列数据(制表符分隔)到下框(第一行为时间(剂量)值, 第二行开始为每个对象观察值, 每行代表一个观察对象)
1. 如有分层变量, 置第一列, 第一行第一列为该变量的变量名。
2. 如有缺失值, 用 NA 表示(无引号)

第一列为分组因素 将分组因素按连续性变量分析

| CONCE | 12 | 24 | 36 | 48 | 72 |
|-------|-------|-------|-------|-------|-------|
| 0 | 99.52 | 99.21 | 98.09 | 99.36 | 98.49 |
| 0 | 99.62 | 99.42 | 99.31 | 99.12 | 99.24 |
| 0 | 98.84 | 98.76 | 99.24 | 99.31 | 99.15 |
| 10 | 96.68 | 92.89 | 86.54 | 76.49 | 65.59 |
| 10 | 95.24 | 91.12 | 85.23 | 77.28 | 65.88 |
| 10 | 95.81 | 90.64 | 84.88 | 76.62 | 64.24 |
| 20 | 89.39 | 76.89 | 69.16 | 64.66 | 51.26 |
| 20 | 89.51 | 75.12 | 68.22 | 62.12 | 50.24 |
| 20 | 88.46 | 75.24 | 68.67 | 62.78 | 50.12 |
| 40 | 76.66 | 67.01 | 55.35 | 50.03 | 44.91 |
| 40 | 73.88 | 68.21 | 55.02 | 48.12 | 42.68 |
| 40 | 75.29 | 66.83 | 56.12 | 48.68 | 44.12 |
| 60 | 68.43 | 56.29 | 46.22 | 38.37 | 33.24 |
| 60 | 66.78 | 54.84 | 44.88 | 37.22 | 32.59 |
| 60 | 68.24 | 55.34 | 45.68 | 36.34 | 33.12 |
| 80 | 59.07 | 41.76 | 34.62 | 26.38 | 10.81 |
| 80 | 58.12 | 39.82 | 35.26 | 24.88 | 10.21 |
| 80 | 58.46 | 41.08 | 34.12 | 24.34 | 11.24 |
| 100 | 42.5 | 27.01 | 18.48 | 7.53 | 3.32 |

拟合时间变量: 按分类变量拟合 一次项加二次项加三次项 一次项加二次项 一次项(直线关系)

结果存到 (默认为我的文档):

上面选择了三种拟合 Y 随 t 的变化趋势：按分类变量拟合，t 加 t^2 两项拟合，只用 t 一次项拟合。点击“开始分析”，即得到如下结果：

结果与解释

数据：略
均数 + 标准差

| Time | All |
|------|-------------------|
| 12 | 75.5214 + 19.7417 |
| 24 | 65.4257 + 24.6383 |
| 36 | 58.1314 + 26.9628 |
| 48 | 51.2276 + 29.6527 |
| 72 | 43.6676 + 30.9816 |

分时间点基本统计

| T | N | Min | Q1 | Median | Q3 | Max |
|----|----|-------|-------|--------|-------|-------|
| 12 | 21 | 42.11 | 59.07 | 75.29 | 95.24 | 99.62 |
| 24 | 21 | 27.01 | 41.76 | 67.01 | 90.64 | 99.42 |
| 36 | 21 | 17.72 | 35.26 | 55.35 | 84.88 | 99.31 |
| 48 | 21 | 7.53 | 26.38 | 48.68 | 76.49 | 99.36 |
| 72 | 21 | 3.18 | 11.24 | 44.12 | 64.24 | 99.24 |

解释：同例 1.1，因为 X 是连续变量，没有按 X 分组统计资料。

随机(截距)效应模型分析时间/剂量-分组因素交互作用：

| Model | Coefficient | Se. | Df. | t-value | P-value | LLR test |
|---------------|-------------|--------|-----|----------|---------|----------|
| Model 1 | | | | | | Ref. |
| (Intercept) | 100.0732 | 2.1368 | 76 | 46.8343 | 0.0000 | |
| factor(t)24 | -4.0982 | 2.1735 | 76 | -1.8855 | 0.0632 | |
| factor(t)36 | -8.8375 | 2.1735 | 76 | -4.0660 | 0.0001 | |
| factor(t)48 | -12.6232 | 2.1735 | 76 | -5.8077 | 0.0000 | |
| factor(t)72 | -19.6461 | 2.1735 | 76 | -9.0389 | 0.0000 | |
| g | -0.5544 | 0.0380 | 19 | -14.5785 | 0.0000 | |
| factor(t)24:g | -0.1354 | 0.0387 | 76 | -3.5010 | 0.0008 | |

| | | | | | | |
|---------------|----------|--------|----|----------|--------|--------|
| factor(t)36:g | -0.1931 | 0.0387 | 76 | -4.9925 | 0.0000 | |
| factor(t)48:g | -0.2635 | 0.0387 | 76 | -6.8126 | 0.0000 | |
| factor(t)72:g | -0.2757 | 0.0387 | 76 | -7.1262 | 0.0000 | |
| Model 2 | | | | | | 0.9148 |
| (Intercept) | 105.0140 | 3.1216 | 80 | 33.6411 | 0.0000 | |
| t | -0.4132 | 0.1511 | 80 | -2.7345 | 0.0077 | |
| t.square | 0.0010 | 0.0017 | 80 | 0.5710 | 0.5696 | |
| g | -0.4125 | 0.0556 | 19 | -7.4252 | 0.0000 | |
| t:g | -0.0135 | 0.0027 | 80 | -5.0338 | 0.0000 | |
| t.square:g | 0.0001 | 0.0000 | 80 | 3.4804 | 0.0008 | |
| Model 3 | | | | | | 0.0000 |
| (Intercept) | 103.6641 | 2.2089 | 82 | 46.9304 | 0.0000 | |
| t | -0.3290 | 0.0400 | 82 | -8.2189 | 0.0000 | |
| g | -0.5590 | 0.0393 | 19 | -14.2186 | 0.0000 | |
| t:g | -0.0044 | 0.0007 | 82 | -6.1777 | 0.0000 | |

对数似然比检验

| Model | Log likelihood | DF. | LRT | LRT |
|-------|----------------|-----|--------|--------|
| 1 | -315.9426 | 12 | Ref. | |
| 2 | -316.4260 | 8 | 0.9148 | Ref. |
| 3 | -333.9853 | 6 | 0.0000 | 0.0000 |

解释：上面运行了三个模型，模型 1 把 t 按分类变量，是全模型。模型 2 用 t 加 t^2 两项拟合 Y 随 t 的变化趋势，模型 3 只用 t 拟合。对数似然比检验结果：模型 2 与模型 1 没有显著差别，而模型 3 与模型 2 及模型 1 相比有显著差别。因此可用模型 2，即 Y 随 t 的变化是二次项的曲线变化，从 t 的回归系数 (-0.4132) 上看 Y 随 t 减速 (t^2 项回归系数 > 0) 的下降。

随机(截距)效应模型：时间/剂量与分组因素独立作用

| Model | Coefficient | Se. | Df. | t-value | P-value | LLR test |
|-------------|-------------|--------|-----|---------|---------|----------|
| Model 4 | | | | | | Ref. |
| (Intercept) | 107.7589 | 1.9659 | 80 | 54.8153 | 0.0000 | |
| factor(t)24 | -10.0957 | 1.8005 | 80 | -5.6073 | 0.0000 | |

| | | | | | | |
|-------------|----------|--------|----|----------|--------|--------|
| factor(t)36 | -17.3900 | 1.8005 | 80 | -9.6587 | 0.0000 | |
| factor(t)48 | -24.2938 | 1.8005 | 80 | -13.4931 | 0.0000 | |
| factor(t)72 | -31.8538 | 1.8005 | 80 | -17.6921 | 0.0000 | |
| g | -0.7279 | 0.0285 | 19 | -25.5242 | 0.0000 | |
| Model 5 | | | | | | 0.9082 |
| (Intercept) | 118.9830 | 2.7388 | 82 | 43.4435 | 0.0000 | |
| t | -1.0126 | 0.1259 | 82 | -8.0422 | 0.0000 | |
| t.square | 0.0058 | 0.0014 | 82 | 3.9768 | 0.0001 | |
| g | -0.7279 | 0.0282 | 19 | -25.7808 | 0.0000 | |
| Model 6 | | | | | | 0.0016 |
| (Intercept) | 111.1474 | 1.9486 | 83 | 57.0410 | 0.0000 | |
| t | -0.5238 | 0.0297 | 83 | -17.6119 | 0.0000 | |
| g | -0.7279 | 0.0281 | 19 | -25.9082 | 0.0000 | |

对数似然比检验

| Model | Log likelihood | DF. | LRT | LRT |
|-------|----------------|-----|--------|-------|
| 4 | -342.6287 | 8 | Ref. | |
| 5 | -342.7251 | 6 | 0.9082 | Ref. |
| 6 | -350.2328 | 5 | 0.0016 | 1e-04 |

解释：模型 4、5、6 对应模型 1、2、3，把 X 与 t 的交互作用项除掉了。

对数似然比检验比较时间/剂量-分组交互作用模型与无交互作用模型

| Model | Log likelihood | DF. | Model | Log likelihood | DF. | LLR test |
|-------|----------------|-----|-------|----------------|-----|----------|
| 1 | -315.9426 | 12 | 4 | -342.6287 | 8 | 0.0000 |
| 2 | -316.4260 | 8 | 5 | -342.7251 | 6 | 0.0000 |
| 3 | -333.9853 | 6 | 6 | -350.2328 | 5 | 0.0000 |

解释：上面是把模型 4 与 1 比较，模型 5 与 2 比较，模型 6 与 3 比较，鉴于模型 2 与 5 可用，我们就看模型 2 中交互作用项是否可以剔除。即看模型 5 与模型 2 相比是否有显著差别，如果有表示 X 与 t 交互作用显著。也就是说，给药浓度不同，细胞活力变化不同。根据模型 1 中 t*X 项（即 t:g）的回归系数（-0.0135）知，浓度（g）越高，Y 下降越快。