

重复测量数据分析实例讲解

陈星霖 陈常中

重复测量指的是：对观察对象某个（类）属性进行多次测量。下面用实例讲解的方法介绍重复测量数据的分析原理与软件操作。

例1.1 三个观察组（0、1、2 分别表示对照组与两个实验组）分别于四个时点（0、3、7、14 小时）测得的肿瘤体积大小数据：

Group	0	3	7	14
0	0.06	0.84	1.71	7.71
0	0.17	1.12	1.96	7.73
.....				
2	0.63	1.02	3.13	3.93

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
.....					
10	95.81	90.64	84.88	76.62	64.24
20	89.39	76.89	69.16	64.66	51.26
.....					

重复测量数据分析 1.1

1、数据格式：

这里观察对象是一个个研究个体（ID），分成两组或多组（组别：X），测量指标是某个变量（Y）的值，测量时点是固定的几个时点（ t_0 、 t_1 、 t_2 、 \dots ）。

例 1.1 三个观察组（0、1、2 分别表示对照组与两个实验组）分别于四个时点（0、3、7、14 小时）测得的肿瘤体积大小数据：

Group	0	3	7	14
0	0.06	0.84	1.71	7.71

0	0.17	1.12	1.96	7.73
0	0.19	1.16	3.24	7.91
0	0.21	1.17	3.46	8.04
0	0.22	1.18	4.44	8.05
0	0.23	1.26	4.56	8.09
0	0.5	1.3	4.84	8.48
0	0.71	1.58	4.93	8.55
0	0.77	1.58	6.11	9.05
0	1.04	1.88	6.19	9.26
1	0.13	0.23	0.75	1.22
1	0.19	0.24	1.3	1.43
1	0.21	0.25	1.48	2.05
1	0.26	0.39	1.58	2.19
1	0.34	0.52	1.59	2.66
1	0.38	0.57	1.69	2.76
1	0.42	0.62	1.79	3.16
1	0.43	0.83	2.39	3.35
1	0.55	0.84	2.52	3.65
1	0.75	0.89	2.84	4.31
2	0.15	0.39	0.8	1.17
2	0.26	0.42	0.9	2.38
2	0.26	0.42	1.18	2.54
2	0.39	0.44	1.44	2.71
2	0.45	0.64	1.8	2.8
2	0.5	0.69	1.9	2.81
2	0.52	0.73	1.93	2.9
2	0.54	0.8	2.78	3.13

2	0.56	0.82	2.94	3.55
2	0.63	1.02	3.13	3.93

这里有三个观察组，用 $X=0、1、2$ 表示；四个时点，用 $t=0、3、7、14$ 天表示，观察指标用 Y 表示。

2、分析目的

首先要明确的是分析目的。对这样的数据，可以做很多比较：如四个时点之间的相互比较，三个分组之间的相互比较，还有在每组内四个时点间的相互比较，每个时点内三个分组间的比较。很多人一上来就开始做这些比较，但有没有想过，为什么要做这些比较？比较出来如有显著性差异又能说明什么？有什么意义？

如果一开始就陷到“组间比较”里面，这样就走偏了。从分析目的角度，这里最重要的是要回答两个问题：

- 1) Y 随 t (时间) 的变化趋势是什么？
- 2) X (分组) 不同， Y 随 t 的变化趋势是否不同？

针对上例数据，分析目的就是 (1) 肿瘤体积随天数的变化是什么样的趋势？
(2) 三组相比这个趋势是否不同？

原来这里强调的是 Y 随 t 的变化“趋势”，不要一下子陷入到某个具体时点。可以想象，如果只做时点之间的比较，假如 $X=1$ 组与 $X=0$ 组相比， $t=7$ 天差异不显著， $t=14$ 天差异显著，那么如果问 $t=10$ 天有没有显著差异呢？这就知道了。那这个结果还有什么多大意义呢？如果分析的是趋势就不一样了，如果结果是 $X=1$ 组比 $X=0$ 组 Y 随 t 增长快，我们就可以说 $t=1$ 时，差别可能还看不出来，但差别就已经有了，当 $t=14$ 时差别就非常明显了。

3、分析思路

明确了分析目的，下面看分析思路：

- 1) Y 随 t 的变化趋势，无外乎：
 - 直线性的：可用 $\beta * t$ 表示
 - 曲线性的： $\beta_1 * t + \beta_2 * t^2$ 表示，或再 $+ \beta_3 * t^3$

- 无规律的（难以用参数来描述的）：把时间当成分类变量，生成哑变量，以表示各时间点相对于第一个时点的变化，如 $\beta_1 * (t=3) + \beta_2 * (t=7) + \beta_3 * (t=14)$

2) Y 随 t 的变化趋势是否有组间差异：

Y 随 t 的变化趋势可以用线图表达，三组就有三条线。如果三组趋势相同，意味着三条线平行，趋势不同也就是三条线不平行。如果变化趋势是上述的直线性的，就是看三个 $\beta * t$ 中的 β 是否有显著差异了；如果是曲线性的，看 $\beta_1 * t + \beta_2 * t^2 + \beta_3 * t^3$ 中的 β_1 、 β_2 、 β_3 ；同理，如果是无规律的，看 $\beta_1 * (t=3) + \beta_2 * (t=7) + \beta_3 * (t=14)$ 中的 β_1 、 β_2 、 β_3 。

4、分析方法

有了分析目的与思路，现在讨论分析方法。

首先假定 Y 随 t 的变化趋势是无规律的，我们生成三个哑变量 t_1 、 t_2 、 t_3 分别表示第 3、7、14 天，即：

当 $t=0$ 时， $t_1=0$ ， $t_2=0$ ， $t_3=0$

当 $t=3$ 时， $t_1=1$ ， $t_2=0$ ， $t_3=0$

当 $t=7$ 时， $t_1=0$ ， $t_2=1$ ， $t_3=0$

当 $t=14$ 时， $t_1=0$ ， $t_2=0$ ， $t_3=1$

先不考虑三个实验组间的差别，建立如下方程拟合数据 Y：

$$Y = \beta_0 + \beta_1 * t_1 + \beta_2 * t_2 + \beta_3 * t_3 + e$$

方程中， i 表示每个体， e 表示每个个体的残差，其均数等于 0。从这个方程中，看出：

- $t=0$ 时， $t_1=0$ ， $t_2=0$ ， $t_3=0$ ，所以 $Y_i = \beta_0 + e_i$ ，因此 β_0 就是 $t=0$ 时的 Y 的均数。
- $t=3$ 时， $t_1=1$ ， $t_2=0$ ， $t_3=0$ ，所以 $Y_i = \beta_0 + \beta_1 + e_i$ ，因此 β_1 就是 $t=3$ 与 $t=0$ 相比的 Y 的均数的差。
- $t=7$ 时， $t_1=0$ ， $t_2=1$ ， $t_3=0$ ，所以 $Y_i = \beta_0 + \beta_2 + e_i$ ，因此 β_2 就是 $t=7$ 与 $t=0$ 相比的 Y 的均数的差。

- $t=14$ 时, $t_1=0$, $t_2=0$, $t_3=1$, 所以 $Y_i = \beta_0 + \beta_3 + e_i$, 因此 β_3 就是 $t=14$ 与 $t=0$ 相比的 Y 的均数的差。

上面这个模型有没有改进的地方呢? 有! 怎么改进呢? 为了理解这个问题, 我们想象一下, 如果我的实验设计是这样的: 不是 30 只小白鼠, 而是 120 只, 分成三组, 每组 40 只, 然后 10 个一组分成四个亚组, 第一个 10 只小白鼠只在 0 天时测了肿瘤体积 (Y), 第二个 10 只小白鼠只在 3 天的时候测 Y , 第三个 10 只小白鼠只在 7 天的时候测 Y , 第四个 10 只小白鼠只在 14 天的时候测 Y 。

这时候, 我们看到每个观察对象只测量了一次, 这样的数据就不是重复测量数据。那它是否适合用上述的模型来分析呢? 答案是肯定的。也就是说, 上面介绍的模型适用于非重复测量的数据, 没有体现重复测量的优势。

重复测量设计的优势是什么呢?

非重复测量设计是在 0、3、7、14 天分别用 10 只不同的小白鼠。不难理解, 如第 3 天的测量结果与第 0 天的差异受两个因素影响, 一是天数变了, 二是测量对象不一样。而如果是重复测量的设计呢, 因为是同一批小白鼠, 可以计算前后的差值, 而这个差值就可以说只是天数 (t) 的变化造成的。这样少了测量对象 (小白鼠) 之间的个体差异。研究效率 (发现阳性结果的概率) 就大大提高了。

如何对上述模型进行改进, 体现重复测量的优势呢?

很简单, 在上述模型 $Y = \beta_0 + \beta_1 * t_1 + \beta_2 * t_2 + \beta_3 * t_3 + e$ 中, 只要我们允许 β_0 是随机的, 即每个个体的 β_0 可以不一样就可以了, 方程改为:

$$Y = \beta_{0i} + \beta_1 * t_1 + \beta_2 * t_2 + \beta_3 * t_3 + e$$

这就是 Mixed (混合) 模型, 混合指的就是随机效应 β_{0i} 与固定效应 β_1 、 β_2 、 β_3 混合在一起。这时候:

- 当 $t=0$ 时, $Y = \beta_{0i}$, 每个个体 i 的 β_{0i} 即为其 $t=0$ 点的测量值。
- 当 $t=3$ 时, $Y = \beta_{0i} + \beta_1 * t_1 + e$, 这样 β_1 就是每个个体 i 的 $t=3$ 点与 $t=0$ 点 Y_i 的差值的平均数。
- 当 $t=7$ 时, $Y = \beta_{0i} + \beta_2 * t_2 + e$, 这样 β_2 就是每个个体 i 的 $t=7$ 点与 $t=0$ 点 Y_i 的差值的平均数。
- 当 $t=14$ 时, $Y = \beta_{0i} + \beta_3 * t_3 + e$, 这样 β_3 就是每个个体 i 的 $t=14$ 点与 $t=0$ 点 Y_i 的差值的平均数。

这里我们看到区别就在于，前面计算的是两个时点“均数的差”，改进后计算的是两个时点“差值的均数”。这等同于一般两组比较的 t 检验与配对的 t 检验，配对 t 检验是 Mixed 模型的一个特例。

如果 Y 随 t 的变化趋势是直线性的，则可以改写 Mixed 模型为：

$$Y = \beta_{0i} + \beta_1 * t + e$$

这时候，当 t=3 时， $Y = \beta_{0i} + 3 * \beta_1 + e$ ；当 t=7 时， $Y_i = \beta_{0i} + 7 * \beta_1 + e_i$ ；依此类推。那么这时候 β_{0i} 是什么？

要理解这个 β_{0i} 是什么，我们可以想象一下，每个个体有四对 Y_{ij} 对应 t_j ，的值，j=0、3、7、14，可以做一个线性回归， $Y_{ij} = \beta_{0i} + \beta_1 * t_j$ ，每个个体的 β_1 相同是固定效应，但 β_{0i} 可以不同，也就是说每个个体的这个方程的截距不同，是随机的。

对照非线性模型，我们不难理解 Mixed 模型：

$$Y = \beta_{0i} + \beta_1 * t + e$$

与一般线性模型： $Y = \beta_0 + \beta_1 * t + e$

两者相比，根本差别在于：Mixed 模型拟合的是每个个体 Y 随 t 的变化斜率的均值，而一般模型拟合的是 Y 在不同时点的均值随 t 变化的斜率。前者在同一个个体内拟合数据，时点前后的个体差异被抵消了。

现在考虑在模型中增加分组因素的作用。这里分组因素 X 有三组，可以生成两个哑变量 X_1 、 X_2 ：

当 X=0 时， $X_1=0$ ， $X_2=0$

当 X=1 时， $X_1=1$ ， $X_2=0$

当 X=2 时， $X_1=0$ ， $X_2=1$

如果 t 也是按哑变量放入模型，则就有这样两种模型：

$$\text{模型 I: } Y = \beta_{0i} + \beta_1 * t_1 + \beta_2 * t_2 + \beta_3 * t_3 + \beta_4 * X_1 + \beta_5 * X_2 + e$$

$$\text{模型 II: } Y = \beta_{0i} + \beta_1 * t_1 + \beta_2 * t_2 + \beta_3 * t_3 + \beta_4 * X_1 + \beta_5 * X_2 + \beta_6 * X_1 * t_1 + \beta_7 * X_1 * t_2 + \beta_8 * X_1 * t_3 + \beta_9 * X_2 * t_1 + \beta_{10} * X_2 * t_2 + \beta_{11} * X_2 * t_3 + e$$

模型 I，没有交互作用项，意味着时点间的差异，在三组中是相同的。

模型 II，有交互作用项，看的是时点间的差异在三组中是否不同。

如 $t=3$ 与 $t=0$ 比（即 $t_1=1, t_2=0, t_3=0$ ）：

- $X=0$ 时（即 $X_1=0, X_2=0$ ），两时点差值是： β_1
- $X=1$ 时（即 $X_1=1, X_2=0$ ），两时点差值是： $\beta_1 + \beta_6$
- $X=2$ 时（即 $X_1=0, X_2=1$ ），两时点差值是： $\beta_1 + \beta_9$

如果 β_6 显著不等于 0，表示 $X=1$ 组与 $X=0$ 组，肿瘤在 $t=3$ 这个点上增长的幅度显著不同。

如果 β_9 显著不等于 0，表示 $X=2$ 组与 $X=0$ 组，肿瘤在 $t=3$ 这个点上增长的幅度显著不同。

余依此类推。

同理，如果 t 是按一次项的连续性变量放入模型，则就有这样两种模型：

模型 I： $Y = \beta_{0i} + \beta_1 * t + e$

模型 II： $Y = \beta_{0i} + \beta_1 * t + \beta_2 * X_1 * t + \beta_3 * X_2 * t + e$

模型 I，没有交互作用项，意味着（ $X=0, 1, 2$ ）三组 Y 随 t 的变化斜率是相同的。

模型 II，有交互作用项，看的是（ $X=0, 1, 2$ ）三组 Y 随 t 的变化斜率是否不同。

- 当 $X=0$ 时（即 $X_1=0, X_2=0$ ）， Y 随 t 的变化斜率是 β_1 ，
- 当 $X=1$ 时（即 $X_1=1, X_2=0$ ）， Y 随 t 的变化斜率是 $\beta_1 + \beta_2$ ，
- 当 $X=2$ 时（即 $X_1=0, X_2=1$ ）， Y 随 t 的变化斜率是 $\beta_1 + \beta_3$ ，

如果 β_2 显著不等于 0，表示 $X=1$ 组与 $X=0$ 组，肿瘤随 t 的变化斜率显著不同。

如果 β_3 显著不等于 0，表示 $X=2$ 组与 $X=0$ 组，肿瘤随 t 的变化斜率显著不同。

5、软件操作

明确了分析目的，确定了分析思路与分析方法，下面介绍如何使用“易侖统计”软件实现上述分析。

从实验室出来的数据通常可以在 Excel 里整理成上表的格式，这里第一列是组别，第一行 Group 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”。

如何拟合时间变量，即如何拟合 Y 随 t 的变化趋势？这里有四个选项，默认选了“按分类变量拟合”与“一次项（直线关系）”两种拟合。按分类变量拟合，即前面分析思路部分所述，不假定 Y 随 t 呈直线或曲线性（含二次项或三次项）变化，按时点 t 生成哑变量放入模型中。如果要看 Y 是否随 t 呈二次项或三次项的曲线性的变化，可以选相应的选项。这里考虑到只有 4 个时点，如果按分类变量处理，只需要三个哑变量即可。如果用 t 加 t^2 拟合就有两个时间参数，而且二次项模型参数解释比较复杂，就不选了。也就是说，这里看能否用线性关系来拟合 Y 随 t 的变化，如果不能，就把 t 按分类变量来拟合。

重复测量数据随机效应模型分析

标题: 输出文件名: 输出路径 (置空表示到我的文档; "."表示到当前工作目录):

复制粘贴数据到下表：

- 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；
- 分层变量置第一列(第一行第一列为其变量名)；缺失数据置空或用NA表示，数据不能含有其它字符或引号。

	A	B	C	D	E	F	G
1	Group	0	3	7	14		
2	0	0.06	0.84	1.71	7.71		
3	0	0.17	1.12	1.96	7.73		
4	0	0.19	1.16	3.24	7.91		
5	0	0.21	1.17	3.46	8.04		
6	0	0.22	1.18	4.44	8.05		
7	0	0.23	1.26	4.56	8.09		
8	0	0.5	1.3	4.84	8.48		
9	0	0.71	1.58	4.93	8.55		
10	0	0.77	1.58	6.11	9.05		
11	0	1.04	1.88	6.19	9.26		
12	1	0.13	0.23	0.75	1.22		
13	1	0.19	0.24	1.3	1.43		
14	1	0.21	0.25	1.48	2.05		
15	1	0.26	0.39	1.58	2.19		
16	1	0.34	0.52	1.59	2.66		

分组变量:

- 第一列为分组因素
 将分组因素按连续性变量分析

拟合时间变量:

- 按分类变量拟合
 $x + x^2 + x^3$

- $x + x^2$
 x (直线拟合)

点击“开始分析”，即得到如下结果：

6、输出结果及解读

数据

Group	0	3	7	14
0	0.06	0.84	1.71	7.71
0	0.17	1.12	1.96	7.73
0	0.19	1.16	3.24	7.91
0	0.21	1.17	3.46	8.04
0	0.22	1.18	4.44	8.05
0	0.23	1.26	4.56	8.09
0	0.5	1.3	4.84	8.48
0	0.71	1.58	4.93	8.55
0	0.77	1.58	6.11	9.05
0	1.04	1.88	6.19	9.26
.....				
2	0.56	0.82	2.94	3.55
2	0.63	1.02	3.13	3.93

解读：这部分是展示首、尾的部分数据，以帮助用户判断数据读入是否有错误。

均数 + 标准差

Time	0	1	2
0	0.410 + 0.3269	0.366 + 0.1860	0.426 + 0.1564
3	1.307 + 0.2961	0.538 + 0.2575	0.637 + 0.2140
7	4.144 + 1.5418	1.793 + 0.6243	1.880 + 0.8377
14	8.287 + 0.5351	2.678 + 0.9784	2.792 + 0.7354

解读：这部分是分组分时点统计Y的均数与标准差

分时间点基本统计

T	Group	N	Min	Q1	Median	Q3	Max
0	0	10	0.06	0.195	0.225	0.6575	1.04
0	1	10	0.13	0.2225	0.36	0.4275	0.75
0	2	10	0.15	0.2925	0.475	0.535	0.63
3	0	10	0.84	1.1625	1.22	1.51	1.88
3	1	10	0.23	0.285	0.545	0.7775	0.89
3	2	10	0.39	0.425	0.665	0.7825	1.02
7	0	10	1.71	3.295	4.5	4.9075	6.19
7	1	10	0.75	1.505	1.64	2.24	2.84
7	2	10	0.8	1.245	1.85	2.5675	3.13
14	0	10	7.71	7.9425	8.07	8.5325	9.26
14	1	10	1.22	2.085	2.71	3.3025	4.31
14	2	10	1.17	2.5825	2.805	3.0725	3.93

解读：这部分是分组分时点更多的基本统计指标：N、最小值、Q1 (25%位数)、中位数、Q3 (75%位数)、最大值。

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)	0.4100	0.2164	81	1.8948	0.0617	
factor(t)3	0.8970	0.2111	81	4.2492	0.0001	
factor(t)7	3.7340	0.2111	81	17.6886	0.0000	
factor(t)14	7.8770	0.2111	81	37.3146	0.0000	
factor(g)1	-0.0440	0.3060	27	-0.1438	0.8867	
factor(g)2	0.0160	0.3060	27	0.0523	0.9587	
factor(t)3:factor(g)1	-0.7250	0.2985	81	-2.4285	0.0174	
factor(t)7:factor(g)1	-2.3070	0.2985	81	-7.7277	0.0000	
factor(t)14:factor(g)1	-5.5650	0.2985	81	-18.6410	0.0000	
factor(t)3:factor(g)2	-0.6860	0.2985	81	-2.2979	0.0241	
factor(t)7:factor(g)2	-2.2800	0.2985	81	-7.6373	0.0000	
factor(t)14:factor(g)2	-5.5110	0.2985	81	-18.4601	0.0000	

Model 2						1e-04
(Intercept)	0.0429	0.1951	87	0.2197	0.8266	
t	0.5824	0.0163	87	35.7637	0.0000	
factor(g)1	0.2423	0.2759	27	0.8781	0.3876	
factor(g)2	0.3137	0.2759	27	1.1367	0.2656	
t:factor(g)1	-0.4059	0.0230	87	-17.6274	0.0000	
t:factor(g)2	-0.4028	0.0230	87	-17.4924	0.0000	

解读：上面列出了两个模型。

模型 1：把 t 按哑变量分析，即 factor(t)3、factor(t)7、factor(t)14 对应 t=3、7、14 三个时点。分组因素(g)是按两个哑变量放入模型，factor(g)1 与 factor(g)2 对应 group=1、2 两组。模型中有 6 个交互作用项。

- Intercept 表示 g=0, t=0 时 Y 的均数，是 0.4100，对照一下前面列出来的分组分时点的均数，完全相同。
- factor(t)3 的回归系数表示 g=0 组，t=3 与 t=0 相比的差，是 0.8970。
- factor(t)7 的回归系数表示 g=0 组，t=7 与 t=0 相比的差，是 3.7340。
- factor(t)14 的回归系数表示 g=0 组，t=14 与 t=0 相比的差，是 7.8770。
- factor(g)1 的回归系数表示 t=0 时，g=1 组与 g=0 组相比的差，是 -0.0440。
- factor(g)2 的回归系数表示 t=0 时，g=2 组与 g=0 组相比的差，是 0.0160。
- factor(t)3:factor(g)1 的回归系数表示 g=1 与 g=0 相比，t=3 与 t=0 两时点差值的差，是 -0.7250。
- factor(t)7:factor(g)1 的回归系数表示 g=1 与 g=0 相比，t=7 与 t=0 两时点差值的差，是 -2.3070。
- factor(t)14:factor(g)1 的回归系数表示 g=1 与 g=0 相比，t=14 与 t=0 两时点差值的差，是 -5.5650。
- factor(t)3:factor(g)2 的回归系数表示 g=2 与 g=0 相比，t=3 与 t=0 两时点差值的差，是 -0.6860。
- factor(t)7:factor(g)2 的回归系数表示 g=2 与 g=0 相比，t=7 与 t=0 两时点差值的差，是 -2.280。

- `factor(t)14:factor(g)2` 的回归系数表示 $g=2$ 与 $g=0$ 相比, $t=14$ 与 $t=0$ 两时点差值的差, 是 -5.5110 。

上面这些结果都可以通过对照前面列出来的均数, 进行核实。

模型 2: 把 t 按连续性变量, 用直线拟合 Y 与 t 的关系。

- t 的回归系数表示 $g=0$ 组 Y 随 t 的变化斜率, 是 0.5824 。
- `t:factor(g)1` 的回归系数表示 $g=1$ 组与 0 组相比, Y 随 t 的变化斜率的差, 是 -0.4059 , $p < 0.0001$ 表示 $g=1$ 组的斜率显著低于 $g=0$ 组。
- `t:factor(g)2` 的回归系数表示 $g=2$ 组与 0 组相比, Y 随 t 的变化斜率的差, 是 -0.4028 , $p < 0.0001$ 表示 $g=2$ 组的斜率显著低于 $g=0$ 组。

模型 2 与模型 1 相比, 参数少很多, 结果简单容易解释。那么是否可以用直线来拟合 Y 随 t 的变化呢? 这就是要比较这两个模型, 结果是 LLR (对数似然比) 检验 $P=0.0001$, 差异非常显著, 也就是说, 简化的模型 2 不足以替代模型 1, 意味着 Y 随 t 的变化不是线性的。

模型 1 与 2 中都只有 $g=1$ 与 $g=0$ 相比, $g=2$ 与 $g=0$ 相比, $g=2$ 与 $g=1$ 比较的结果有没有呢? 没有! 易俩自动把其它各组与 0 组比较, 如果要比较 $g=2$ 与 $g=1$, 可以把 `group` 的取值重新编码一下, 把 $g=1$ 改成 $g=0$ 即可得到。

对数似然比检验

Model	Log likelihood	DF.	LRT
1	-99.1760	14	Ref.
2	-113.7268	8	1e-04

解读: 这是上面列出了两个模型对数似然比检验结果, 模型 1 作为参照。模型 1 有 14 个自由度, 模型 2 的只有 8 个, 但模型 2 得出的对数似然值比模型 1 小很多。

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 3						Ref.

(Intercept)	1.8328	0.2746	87	6.6739	0.0000	
factor(t)3	0.4267	0.3171	87	1.3455	0.1820	
factor(t)7	2.2050	0.3171	87	6.9534	0.0000	
factor(t)14	4.1850	0.3171	87	13.1973	0.0000	
factor(g)1	-2.1932	0.2746	27	-7.9863	0.0000	
factor(g)2	-2.1033	0.2746	27	-7.6586	0.0000	
Model 4						0.1554
(Intercept)	1.6604	0.2343	89	7.0857	0.0000	
T	0.3128	0.0215	89	14.5301	0.0000	
factor(g)1	-2.1933	0.2765	27	-7.9321	0.0000	
factor(g)2	-2.1033	0.2765	27	-7.6066	0.0000	

解读：这上面列出了模型 3 与模型 4，分别对应于前面的模型 1 与模型 2，只是把交互作用项除掉了。

对数似然比检验

Model	Log likelihood	DF.	LRT
3	-191.8573	8	Ref.
4	-193.7187	6	0.1554

解读：这上面是模型 3 与模型 4 对数似然比结果。P=0.1554，差异没有显著性。也就是说，如果不考虑分组 g 与 t 的交互作用，是可以直接用直线来拟合 Y 随 t 的变化。

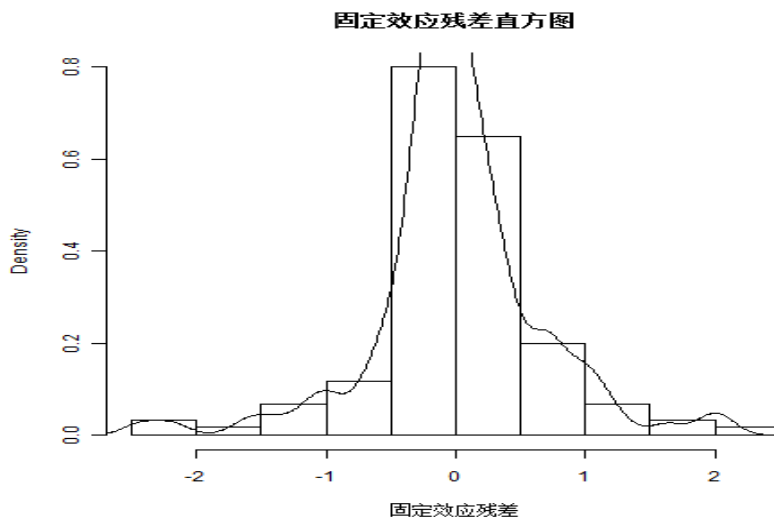
对数似然比检验比较时间/剂量-分组交互作用模型与无交互作用模型

Model	Log likelihood	DF.	Model	Log likelihood	DF.	LLR test
1	-99.1760	14	3	-191.8573	8	0.0000
2	-113.7268	8	4	-193.7187	6	0.0000

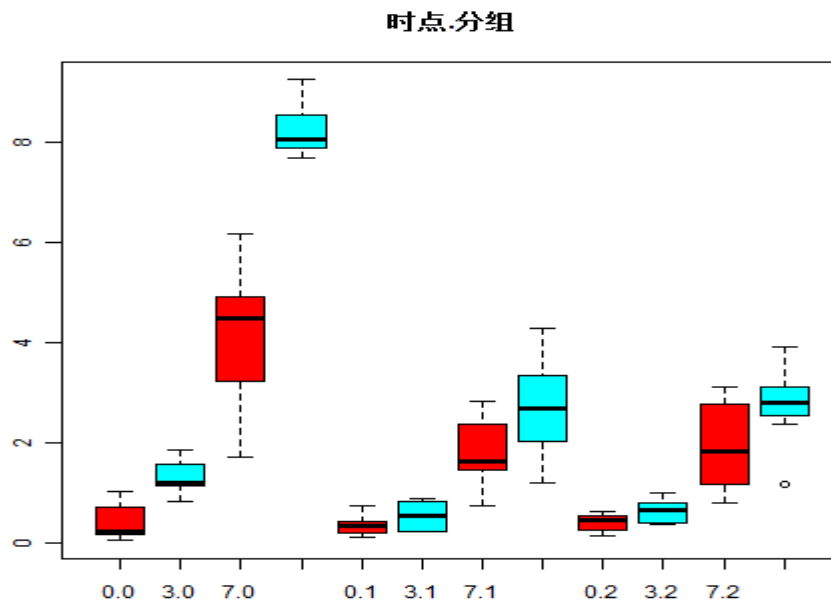
解读：这上面是把模型 3 与模型 1 相比，把模型 4 与模型 2 相比，进行对数似然比检验，结果是差异均有显著性。也就是说，如果不管怎么拟合 Y 随 t 的关系，分组 g 与 t 的交互作用都是显著的。

	统计量(w)	p 值	检验方法	备注
正态性检验	0.9180603	7.79313e-06	Shapiro-Francia	固定效应残差分析

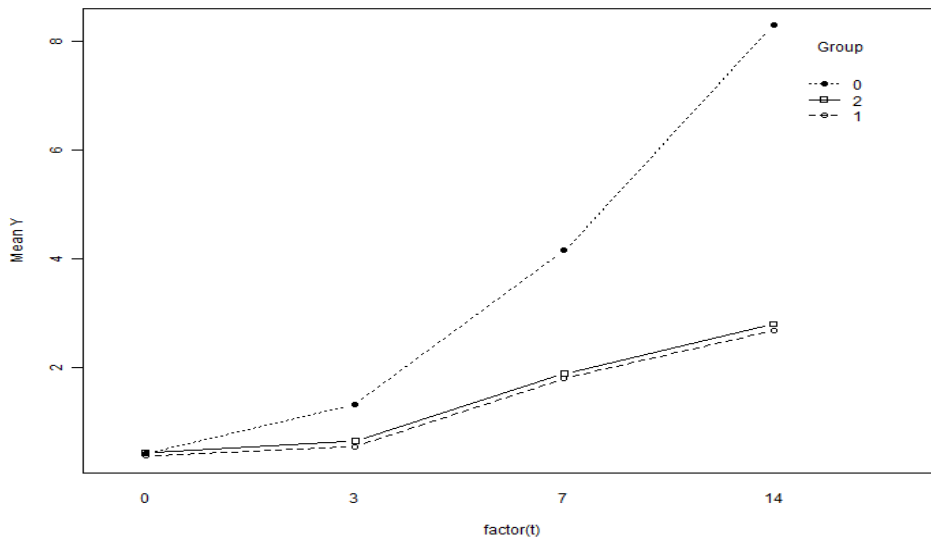
解读：这上面是对固定效应的残差进行正态性检验。这里正态性检验结果 $p=7.793 \times 10^{-6}$ ，非常显著，即不符合正态性。这个结果仅供参考，如果是正态最好，如果不符合，要看下面的分布图，如果基本对称，仍然是可以用上述分析的。



解读：这是固定效应残差的直方图，可以看出基本上左右对称。



解读：这是按时点按分组的箱图，基本看出（1）各组 Y 的分布呈上下对称；
（2）各组分时点的差异。



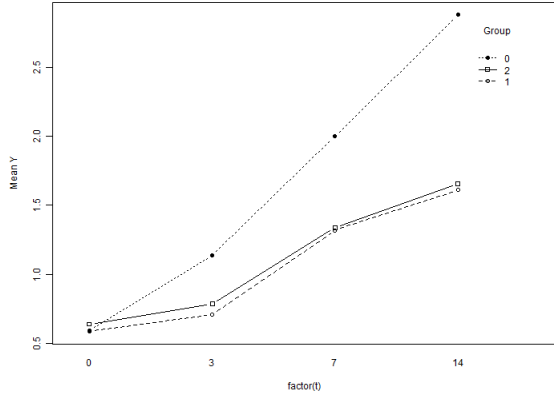
解读：这是分组的 Y 随 t 的变化线图，很明显 g=0 组与 1、2 组 Y 的变化斜率不同。也看出 Y 随 t 的变化趋势是加速上升的。

补充分析：

根据上述分析结果，我们看到（1）固定效应残差非正态，（2）Y 随 t 有加速上升的趋势。可以考虑把数据进行对数或开方转换，再重新分析。下面是把原始数据开方后，正态性检验与图形结果：

	统计量(w)	p 值	检验方法	备注
正态性检验	0.9784744	0.04981288	Shapiro-Francia	固定效应残差分析

可以看到正态性检验的 P 值接近 0.05 了。图形结果如下：



这个图显示 Y 随 t 的变化基本上可以认为是直线性的上升了。

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“易侓统计”，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

标题:

重复测量数据分析

输出文件名:

mixed

输出路径 (置空表示到我的文档; "."表示到当前工作目录):

复制粘贴数据到下表:

- 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;
- 分层变量置第一列(第一行第一列为其变量名); 缺失数据置空或用NA表示, 数据不能含有其它字符或引号。

Clear table

Example data

开始计算

	A	B	C	D	E	F	G
1	CONCE	12	24	36	48	72	
2	0	99.52	99.21	98.09	99.36	98.49	
3	0	99.62	99.42	99.31	99.12	99.24	
4	0	98.84	98.76	99.24	99.31	99.15	
5	10	96.68	92.89	86.54	76.49	65.59	
6	10	95.24	91.12	85.23	77.28	65.88	
7	10	95.81	90.64	84.88	76.62	64.24	
8	20	89.39	76.89	69.16	64.66	51.26	
9	20	89.51	75.12	68.22	62.12	50.24	
10	20	88.46	75.24	68.67	62.78	50.12	
11	40	76.66	67.01	55.35	50.03	44.91	
12	40	73.88	68.21	55.02	48.12	42.68	
13	40	75.29	66.83	56.12	48.68	44.12	
14	60	68.43	56.29	46.22	38.37	33.24	
15	60	66.78	54.84	44.88	37.22	32.59	
16	60	68.24	55.34	45.68	36.34	33.12	

分组变量:

第一列为分组因素

将分组因素按连续性变量分析

拟合时间变量:

按分类变量拟合

$x + x^2 + x^3$

$x + x^2$

x (直线拟合)

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t²两项拟合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(t²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。