混合效应模型

2021年5月12日 14:33

- ## 基本概念和特征：

✓ 在fisher的ANOVA中是把所有的分类解释变量视为相同的，而Eisenhart认识到了实际上有两种不同的分类解释变量：固定效应和随机效应。固定模型：将结果概括为研究中使用的实验值，例如药物，男性，女性。随机模型：做出了超出研究的特定值的推论，例如地点和个人

✓ 混合模型：固定效应仅影响y的平均值；随机效应仅影响y的方差。

✓ 在混合效应模型中，随机效应来自一个很大的总体，因此没有必要集中精力估计我们的一小部分因素水平的平均数，也没有必要比较不同因素水平的单个平均数对。更好的做法是从更大的总体中随机抽取样本并且关注它们的方差。因此随机效应模型类似单因素方差分析：（其中α和ε的平均值为零，但方差分别为S2 a和S2ε。）

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad i = 1,\ldots, \text{a levels}; j = 1,\ldots, \text{n obs.}$$

where the α and ε have mean zero, but variances σ^2_α and σ^2_ε respectively.

$$\sigma^2_\varepsilon = SSE/(a(n-1)) = MSE$$

$$\sigma^2_\alpha = (MSA - MSE) / n$$

- 因子与协变量的区别：
  - 首先，两者都是自变量。区别在于测量水平：前者是名目或定类变量（只含两至数个类别，每个类别中至少要有30个案例），后者是连续或定距变量（可以含成千上百个类别，每个类别中只含一至数个案例）。当你通过这种区别、将每个自变量的测量水平告知SPSS或其它统计软件之后，软件就知道如何计算了。这是测量层面(operationalization)上讨论“因子”和“协变量”之间的区别。
  - 它们在理论(conceptualization)上的含义很不同，不要混淆：因子可以是自变量（如外生因子）、也可以说因变量（如内生因子），两者即可以是名目变量、也可以是连续变量；协变量则被用来指“控制变量”（也是对因变量有影响的自变量、但不是理论上所关心的，所以引入以过滤其影响），可以是连续变量（如年龄）、也可以是名目变量（如性别）

- 固定变量与随机变量之区别:
  - 在GLM范畴内，所谓“固定”和“随机”变量，表面上是指自变量，其实是涉及数据结构。什么意思？一、你的因变量有几个？二、你的自变量之间是“同级并列”（如个人特征）还是“上下主从”（如个人特征在下、机构或社区特征在上）？这两个维度相交，形成了下表中的四种情况。

	自变量之间关系	
因变量个数	同级并列	上下主从
单个	I. 固定	II. 随机
多个	III. 随机	IV. 随机

  - I. 单个因变量和并列自变量：这是最常见（但也是最有局限）的数据，自变量必定是固定的。
  - II. 单个因变量和主从自变量：由于自变量之间有从属关系，所以形成了multilevel(多层)结构。为了与III和IV相区别，我将这种多层结构叫做“variances between-subjects”（BS差异或人际差异）。下层的自变量为随机而上层的自变量为固定。如果有3+层的话，最高一层为固定、以下各层均为随机。

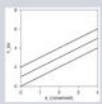
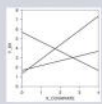
- III. 多个因变量和并列自变量：这时，虽然自变量之间是并列的，但因变量之间存在着明显（如同一概念时间先后的测量）或隐含（同一大概念下面的众多子概念）的关系，所以形成了与BS差异不同的另一种多层结构，我称之为variances within-subjects（WS差异或自身差异）。这种结构中，多个因变量的序号构成了下层自变量的值；而其上层自变量并不存在，需要在GLM或类似程序中构建相应的虚拟变量（我知道，这句话对没有实际操作经验者来说是很难懂的，如不理解就暂且跳过吧）。但是，WS差异结构与BS差异结构相同的是，最高层的自变量总是固定的，而以下各层的自变量均是随机的。
- IV. 多个因变量和主从自变量：这种结构同时兼有BS差异和WS差异（即BS-WS差异），其最上层的BS自变量和WS自变量都是固定的而以下各层各种变量都是随机的。这当然是最丰富、也是最难得的数据，如固定样本数据的SEM模型。
- 需要注意的是，随机变量可以当作固定变量处理（当然有犯Type-I错误，即可能过高估计自变量的影响。实际研究上，很多II类数据被当作I类处理）。但是反之不亦然，固定变量不可能成为随机变量。

- 随机效应模型是所有因子都表示随机效应的模型。（请参见随机效应。）此类模型亦称方差分量模型。随机效应模型往往是分层模型。同时包含固定和随机效应的模型称为混合模型。重复测量和裂区模型是混合模型的特例。通常使用混合模型一词来包含随机效应模型。固定效应模型和随机效应模型区别如何理解固定效应模型： 固定效应模型和随机效应模型的区别在于其基本假设不同。 前者认为效应是固定的，且误差项和解释变量相关；后者认为效应是随机的，误差项和解释变量不相关。 因此固定效应模型更适用于研究样本之间的区别，而随机效应更适用于由样本来推断总体特征。固定效应（fixed effect, FE）vs. 随机效应（random effect, RE）
- **随机效应**
  随机效应是其水平被视为某个总体中的随机样本的因子。通常，随机效应的精确水平并不受关注，而水平反映出的变异才受关注（方差分量）。不过，也有些时候您想要预测随机效应给定水平的响应。更严格地来讲，随机效应被视为服从均值为零且方差非零的正态分布。
  假定您关注的是两种特定的烤炉对模具收缩的影响是否不同。一个烤炉一次仅处理一批 50 个模具。您设计这样一个研究方案：先后分别为每一个烤炉随机选三个批次（每批 50 个模具）放进烤炉。这些批次的模具处理完毕后，对每批中随机选择的五个部件测量收缩。
  请注意，“批次”是包含六个水平的因子，每批次对应一个水平。所以，在您的模型中包括两个因子 —“烤炉”和“批次”。由于您特别关注比较每个烤炉对于收缩的影响，所以“烤炉”是一个固定效应。但您并不关注这六个特定批次对于收缩均值的影响。这些批次代表了一个都有可能被这个实验选上的批次总体，而且其分析结果必须推广到的整个批次总体。“批次”被视为随机效应。在该实验中，“批次”因子因为在所有可能批次中的收缩变异而受到关注。您想要估计它所解释的收缩变异量。（请注意，“批次”还嵌套在“烤炉”中，因为在一个烤炉中一次只能处理一个批次。）

- **“混合模型”**特质允许您分析具有复杂协方差结构的模型。可以分析的情形包括：
  - 裂区实验：裂区实验是包含实验单元的两个或更多水平（或大小）而导致多个误差项的实验。当一些因子易于改变而另一些因子难以改变时，通常需要进行这样的设计
  - 随机系数模型：随机系数模型也称为层次或多水平模型 (Singer 1998；Sullivan et al. 1999)。当认为批次或对象的截距和斜率随机变化时，使用这些模型。制药行业的药物稳定性试验和教育研究领域的个人成长研究通常需要随机系数模型。
  - 重复测量设计：重复测量设计也称为对象内设计，它对响应在时间或空间上的变化建模并允许误差有相关性。
  - 空间数据：空间数据是在二维或多维中进行的测量，通常为纬度和经度。空间测量值的相关性通常表示为其空间邻近度的函数。
  - 相关响应数据：相关响应数据来自对同一实验单元的几次测量。例如，医学研究中对个体所测的身高、体重和血压读数可能是相关的，制造业中对产品所测的硬度、强度和弹性值可能是相关的。尽管可以单独研究这些测量值，将它们作为相关响应处理可以得到有用的信息。
    - 重复测量设计、空间数据和相关响应数据都具有观测值不独立的特点，要求您对相关性结构进行建模。
    - 未考虑观测值之间的相关性可能导致得到有关处理效应的不正确结论。但是，估计协方差结构参数需要

使用数据中的信息。要估计的参数数目影响功效和第一类错误率。因此，您必须正确选择协方差模型

统计学中的「固定效应 vs. 随机效应」

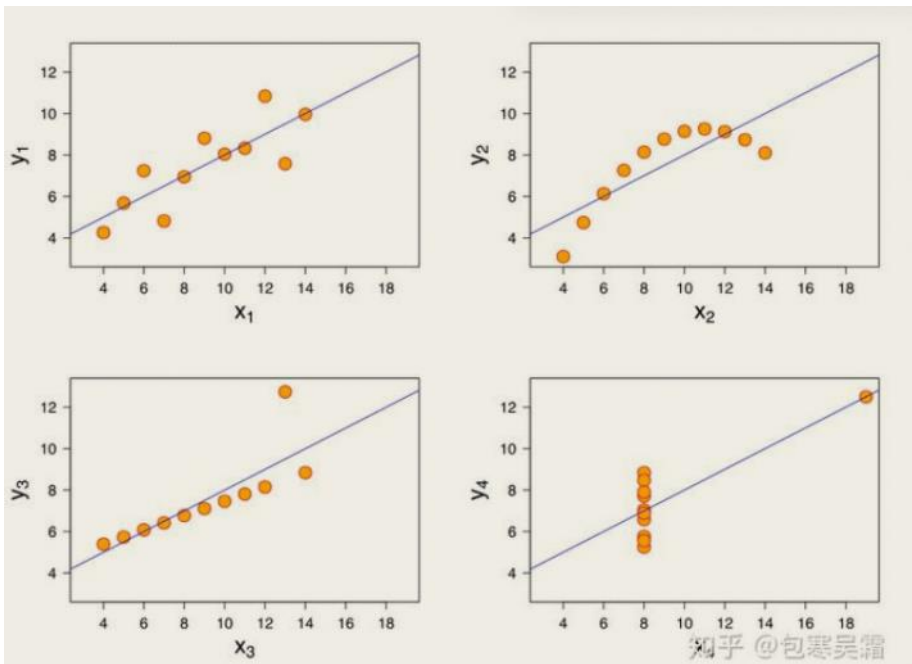
	固定效应 (Fixed Effect, FE)	随机效应 (Random Effect, RE)
总框架：回归分析	$Y = b_0 + b_1*X_1 + b_2*X_2 + \ldots + \text{error}$（· 个体差异 ε_{ij} · 群体差异 u_j） [观测项] = [结构项]（固定部分） + [误差项]（随机部分）	
多层线性模型 (HLM) * 又称：线性混合模型（LMM） * 两大应用情形： – 个体嵌套于群体、组别（横截面数据） – 重复测量嵌套于个体（纵向追踪数据）	**固定截距**（非HLM）：Y ~ 1 + X – 实质：GLM（降级为OLS回归）	**随机截距**：Y ~ 1 + X + (1 | group) – 每组有不同"基线"且正态
	固定斜率：Y ~ 1 + X + (1 | group) – $X_{Lv.1}$效应各组一致 – 组数过少时$df_{Lv.2}$小，建议使用固定斜率 	**随机斜率**：Y ~ 1 + X + (X | group) – $X_{Lv.1}$效应依组而变 – 有跨层交互作用时需要使用随机斜率 
	* 结果中的"固定效应"是指回归系数（"平均"的截距和斜率）	* 结果中的"随机效应"是指残差方差（不同组、不同个体的"特异"部分）
方差分析 (ANOVA) * 可视为GLM和HLM的特例	**固定因素 / 组别** – 实质：固定截距（in GLM） – 组别：$k-1$个虚拟编码 – 组别已涵盖总体的所有取值（仅希望推论到已有组别） – 多数ANOVA都为固定效应（即总体均值固定）	**随机因素 / 组别** – 实质：随机截距（in HLM） – 组别：残差服从正态分布 – 组别仅为总体的随机抽样（希望推论到更大范围） – HLM的零模型等同于"随机效应单因素ANOVA"
面板数据模型 (Panel Model, PLM) * 可视为GLM和HLM的扩展 – $N \times T$ 纵向追踪数据 – "个体"效应 i & 时间效应 t（"个体" = 人、省市……） – 计量经济学的常用模型	**固定效应模型**（在HLM中无对应） – 实质：异质性(非随机)截距（对"个体"生成$N-1$个虚拟变量） – 估计方法：虚拟变量OLS回归（LSDV）、个体内均值离差法、一阶差分法（FD）	**随机效应模型**（HLM的特例） – 实质：异质性(随机)截距（残差需服从正态分布且独立） – 估计方法：可行广义最小二乘法（FGLS）、最大似然法（ML）
	Hausman检验 — H_0：u_i与X_{it}不相关，采用RE；H_1：u_i与X_{it}相关，采用FE	
	▲ 更多采用固定效应模型（FE） — 个体固定 / 时间固定 / 双向固定	
元分析 (Meta-Analysis) * HLM的特例 – 只有Level 2（组间）模型 – Study的效应量和误差实为HLM Level 1的斜率和残差	**固定效应模型** – 实质：固定"斜率"（in HLM）（效应量之间不存在异质性） – 估计方法：Mantel-Haenszel	**随机效应模型 & 元回归** – 实质：随机"斜率"（in HLM）（效应量之间存在异质性） – 估计方法：ML、REML……
	元分析的"截距" ⇔ HLM Level 1的"斜率"；元回归相当于检验跨层交互作用	
	▲ 更多采用随机效应模型（RE），参数估计更保守，也更符合实际情况	

- 在回归分析中：
  - 世界很复杂，但在科学的视角下，很多现象或行为都可以被简化为回归模型——现象或行为本身是“观测项”，我们还会用一系列其他变量来解释或预测这个观测项，其中，一部分是我们能够预测的“结构项”，另一部分则是我们暂时无法预测的“误差项”。
    - 结构项是由一些变量（X1、X2...）及参数（b0、b1...）组成的，尽管我们无法穷尽所有可能的预测变量，但至少我们可以从已知的变量关系中发现一些规律，于是结构项就构成了回归模型的“固定部分”。
    - 误差项则是我们为了简化模型而不得不舍弃的一部分，这种刻意的忽略是不可避免的，否则就会造成“过拟合”（overfitting）。具体来说，误差项又有三个来源：遗漏的变量、测量的误差、随机的干扰。但无论如何，我们终究要在“精确性”（accuracy）和“简约性”（parsimony）之间做出权衡，从而舍弃一部分信息，这些“剩下来”的未被解释的信息就构成了回归模型的“随机部分”。

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + error$$

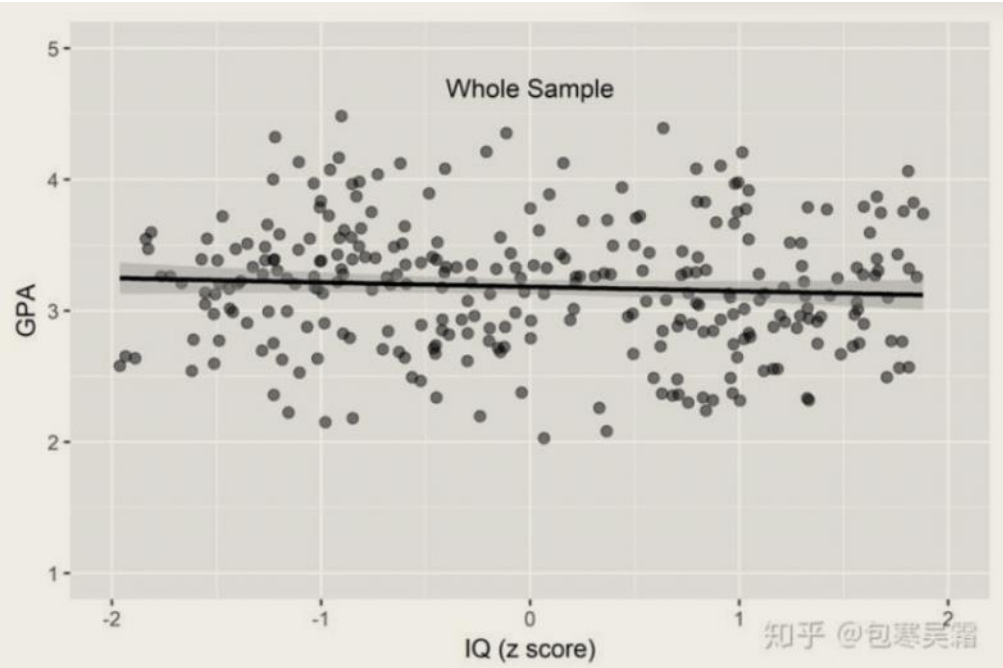
$$[\text{观测项}] = [\text{结构项}]_{(\text{固定部分})} + [\text{误差项}]_{(\text{随机部分})}$$

  - 下面的四幅图直观地展示了回归模型的“固定”与“随机”。同样都是一条回归方程（有几乎相同的截距和斜率，即模型的“固定部分”），但数据的实质却截然不同
    - ①中的X和Y是两个正态分布的变量，其回归模型的“随机部分”基本都来自于随机误差，因此模型是适当的；
    - ②中的X和Y实则是非线性关系，因此用一般的线性回归做拟合是错误的，应加入X的二次项做多项式回归；
    - ③中的一个数据点成为了异常值（outlier），同样会影响回归模型的准确性，可以剔除该点，或者做稳健回归；
    - ④进一步告诉我们，哪怕是一个小小的异常数据点，也足以产生错误的、有误导性的结果。
    - ②~④的共性在于，残差并不满足正态分布，或者存在异方差（heteroscedasticity），所以它们得到的回归模型（固定部分）都是不妥当的。一般而言，回归模型的“随机部分”需要尽可能服从正态分布，这样才能保证“固定部分”的参数估计是无偏的、一致的、有效的。

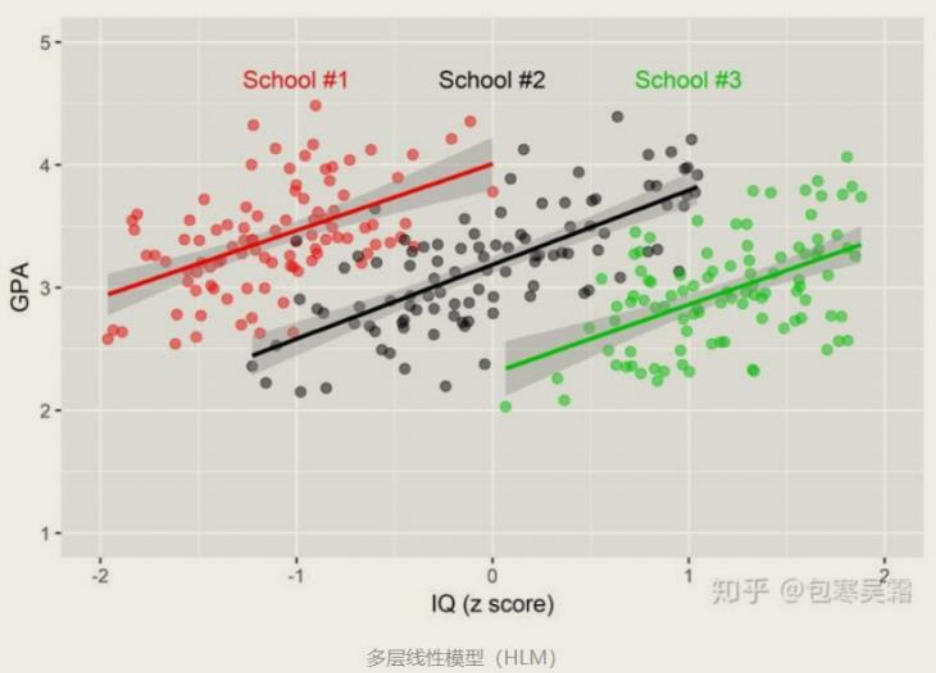


回归模型的恰当性（著名的Anscombe四组数据）

- ## 多层线性模型（HLM）：
  - “HLM是众多统计方法的「幕后主谋」。”什么意思呢？一般我们收集完数据之后会做什么分析？做实验的人会说：方差分析（ANOVA）；做调查的人会说：多元回归（multivariate regression）；做文献综述的人会说：元分析（meta-analysis）……简而言之：回归分析是几乎所有统计模型的基础，而回归分析的最一般形式则可以归为多层线性模型HLM，其他的都可以视为HLM的特例：
    - t检验是ANOVA的一个特例（自变量只有两水平）
    - ANOVA是回归分析的一个特例（自变量为分类变量）
    - 回归分析的实质是一般线性模型GLM（其推广则是广义线性模型）
    - GLM是HLM的一个特例（只有Level 1）
    - 元分析可以视为只有组间模型的HLM（只有Level 2）
  - 当然，除去上面这些HLM的特例，HLM本身关注的焦点其实是“多层嵌套数据”。举个例子：假设我们想知道“智力水平（IQ）能否影响学业成绩（GPA）”，然后我们收集了很多很多学校的数据，迫不及待地画了一个散点图，如图一：这样一看，IQ和GPA并无什么关系呀，甚至还有一点负相关？！于是我们就下结论“IQ与GPA无关”吗？emmm等等，我们漏了什么？事实上，不同学校之间在很多方面可能都存在固有的差异，比如学生整体的智力情况、教学条件、师资力量等等。现在考虑一个情形：收集的学校里面既有“精英学校”也有“普通学校”，学生入学的标准就是他们的IQ。然后我们重新对不同学校分别画一下散点图，看图二，图二中就是经典的“组内同质、组间异质”的情况，我们刚刚做的其实就可以视为HLM的其中一种子模型——随机截距-固定斜率模型，也就是假定不同学校的基线水平不同（随机截距），但IQ与GPA之间的变量关系在不同学校中保持相同（固定斜率）。



一般线性模型（GLM），即最普通的回归分析



多层线性模型（HLM）

- HLM会把多层嵌套结构数据在因变量上的总方差进行分解：总方差 = 组内方差（Level 1）+ 组间方差（Level 2）。比如在上面的例子中，学生是个体水平（Level 1）的分析单元，IQ和GPA都是在个体水平收集的变量，而学校是群体水平（Level 2）的分析单元，不过我们暂时并没有收集学校水平的任何自变量，只是把学校本身当做一个分组变量（clustering/grouping variable）。换句话说，上面这个例子也可用被称作“随机效应单因素协方差分析（ANCOVA with random effects）”。现在用公式来表示上面这个例子：

$$\text{Level 1:}\quad (GPA)_{ij} = \beta_{0j} + \beta_{1j}(IQ)_{ij} + \varepsilon_{ij} \qquad Var(\varepsilon_{ij}) = \sigma^2 \text{ (组内方差[残差])}$$
$$\text{Level 2:}\quad \beta_{0j} = \gamma_{00} + u_{0j} \text{ (随机截距)} \qquad Var(u_{0j}) = \tau_{00} \text{ (组间方差)}$$
$$\beta_{1j} = \gamma_{10} \text{ (固定斜率)}$$

```
# R语言 - 固定斜率:
model = lmer(GPA ~ IQ + (1|School), data=data)
```

- 即理论上假定IQ与GPA之间的关系在不同学校是不一样的，可以设置IQ为随机斜率，

```
# R语言 - 随机斜率:
model = lmer(GPA ~ IQ + (IQ|School), data=data)
```

- 当然，我们还可以引入学校水平的自变量来对学校间的GPA均值差异进行解释，比如教师数量、教学经费……这些变量由于只在学校层面变化，对于每个学校内的每一个学生而言都只有一种可能的取值，因此必须放在Level 2的方程中作为群体水平自变量，而不能简单地处理为个体水平自变量——这也就是HLM的另一个存在的意义：可以同时纳入分析个体与群体水平的自变量。

我们来看HLM的一般式：

Level 1（组内/个体水平，或重复测量/追踪设计中的时间水平；p个自变量，i个样本量）：

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \ldots + \beta_{pj}X_{pij} + \varepsilon_{ij}$$

Level 2（组间/群体水平，或重复测量/追踪设计中的个体水平；q个自变量，j个样本量）：

$$\beta_{pj} = \gamma_{p0} + \gamma_{p1}W_{1j} + \gamma_{p2}W_{2j} + \ldots + \gamma_{pq}W_{qj} + u_{pj}$$

- 也有不少同学把“多层线性模型”与“分层/逐步多元回归”搞混淆——请注意：多层线性模型HLM解决的是多层嵌套结构数据（落脚点是数据结构）。分层/逐步多元回归本身是普通的回归分析，解决的是不同自变量的

重要性的优先程度（落脚点是变量重要性）

- 截距与斜率：
  - **固定截距（fixed intercept）**：固定截距其实并不存在于HLM的模型中，而是“降级”到了一般的最小二乘法回归（OLS），也就是我们最常用的GLM回归分析。
    → lm(Y ~ 1 + X1 + X2, ...)
  - **随机截距（random intercept）**：在做HLM时，我们通常都会将截距设置为随机截距，也就是允许不同组具有各自的截距（基线水平）。可以理解为，“有的人出生就在终点，而你却在起点”。在R里面，只要你在回归表达式后面加上小括号（当然，这时就不能再用lm了，要用lme4和lmerTest包的lmer函数），括号里就定义了Level 1截距或斜率在Level 2的随机部分（Level 1的随机部分则是个体层面的残差residual，不用我们定义）。竖线“|”后面是分组变量（clustering/grouping variable，可以是省市、学校，而在重复测量、追踪设计中则是被试个体），竖线前面的1代表随机截距、具体变量名则代表这个变量对应的随机斜率。
    → lmer(Y ~ 1 + X1 + X2 + (1 | group), ...)
  - **固定斜率（fixed slope）**：固定斜率的意思是，某个Level 1自变量的斜率在不同的group里面都是一致的。虽然实际情况未必真的一致，但研究者可以假设并检验斜率是否在组间保持一致而不存在显著差异。需要注意的是，Level 2截距或斜率并不存在固定和随机的区分（或者说都是固定的），除非还有Level 3。
    → lmer(Y ~ 1 + X1 + X2 + (1 | group), ...)
  - **随机斜率（random slope）**：与固定斜率相反，随机斜率意味着某个Level 1自变量的斜率在不同的group之间存在差异，或者说“依组而变”。可以理解为，“有的人花两个小时就能赚10000，而你却只能挣个10块钱被试费”。你既可以只纳入随机斜率成分而不对斜率的差异作出具体解释，也可以再纳入一个Level 2的自变量与这个Level 1自变量发生交互作用（即跨层交互），从而解释为什么X的效应依组而变、是什么因素导致了这种变化。
    → lmer(Y ~ 1 + X1 + X2 + (1 + X1 | group), ...) # 这里的1可以省略，默认都纳入截距（但只有随机截距时则不能省）
    → lmer(Y ~ 1 + X1*W + X2 + (1 + X1 | group), ...) # W表示一个Level 2解释变量，X1*W即为一个跨层交互作用

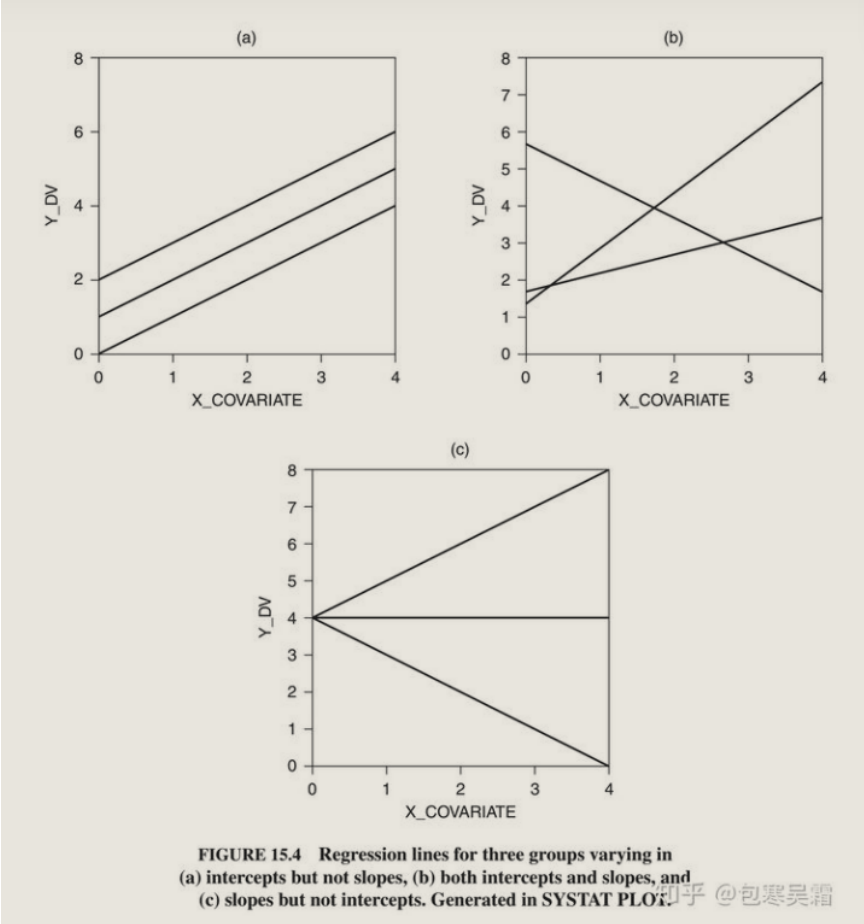


FIGURE 15.4 Regression lines for three groups varying in (a) intercepts but not slopes, (b) both intercepts and slopes, and (c) slopes but not intercepts. Generated in SYSTAT PLOT.

知乎 @包寒吴霜

- 斜率选固定还是随机？（一般来说，只要使用HLM，都会考虑随机截距）
  - 理论假设：我们一般不会特意假设存在随机斜率，也就是说，设置随机斜率需要有比较强的理论假设认为Level 1自变量的效应依组而变。这一点也可以进行统计检验，如果斜率的随机部分（方差成分）与零没有显著差异，则可以考虑舍弃随机斜率。
  - 研究目的：如果我们的研究目的就是想检验是否存在跨层交互作用，那么相应的Level 1自变量就需要设置为随机斜率。这一点容易理解，因为跨层交互本身就意味着X的斜率会变、并且我们还想解释为什么会变，如果你不设置为随机斜率，岂不自相矛盾？
  - Level 2的组数量：如果Level 2的组数量过少（如<10组），则Level 2的自由度过小，不足以做出稳健的参数估计，因此更适合用固定斜率。——BUT！如果Level 2的组数量真的小于10，我们甚至需要仔细斟酌一下是否还要用HLM！因为Level 2的组别也是一种随机取样，在GLM里面我们一般要求样本量至少是变量数的10倍，在HLM里面我们同样会要求组别不能太少，一般要求Level 2至少达到10~20组以上，否则Level 2分模型的统计检验力power就会严重不足！这一点和GLM是一样的！——所以，如果没有10组以上的取样，最好还是不要用HLM，此时我们要改用所谓的“固定截距”模型，即降级到GLM / ANOVA / ANCOVA！
  - Level 2每组内的样本量：联想到经典的被试间设计ANOVA，我们一般要求每组样本量不能太少，至少要30~50名被试才能保证power。这一点同样适用于HLM。如果Level 2每组内的样本量都过小（如<30），虽然仍旧可以使用HLM，但为了得到更稳健的参数估计，需要借助更多乃至所有的数据信息，因此更适合用随机斜率。

    注：在我的上一篇文章《多层线性模型（HLM）及其自由度问题》中，举了一个“三所学校GPA ~ IQ”的例子来说明做HLM的必要性，但实际上，如果真的遇到只有三所学校的情况，我们还真的不太能做HLM，因为组数量太少了！对于这个例子，最合适的做法是不用HLM，而是用协方差分析ANCOVA，把学校作为一个“固定效应”控制掉即可（也相当于为三所学校生成了两个虚拟变量，做虚拟变量OLS回归，等价于ANCOVA）。

- HLM的自由度
  - 自由度的定义：
    - 自由度（degree of freedom, df），简单来说就是“能够独立变化的数据量”。比如一组有N个数据的样本，其总体平均值是确定的，那么在参数估计中，我们说它的自由度是N – 1，这里的1代表已经确定了的均值，也就是说无论你的数据怎么变，只要给我N – 1个数，我就能知道剩下的那个数是几，因为均值已经确定了。
    - 而在普通的回归分析中，同样道理，截距相当于一个已经确定了的均值，是我们要估计的一个参数，并且每增加一个自变量，都会相应增加一个回归系数，这些回归系数也是我们要估计的参数，也视为确定值，因此剩下的那些不确定的、能够独立变化的数据的个数就是自由度，主要用于假设检验（对回归系数的检验：t = b/SE，服从df = N – k的t分布，其中N为总样本量，k为所有的参数个数，截距也占一个参数）。
    - > df（能够独立变化的数据量）= N（总数据量）– k（待估计的参数个数 [包括截距]）
      >
      > 事实上任何一个回归方程都有且仅有一个截距，就好比任何一组数据都有且仅有一个平均值，所以很多时候上面这个表达式也可以用“自变量/预测变量”的个数来表示：
      >
      > $df = N - k = N - p - 1$（p表示自变量/预测变量的个数，1表示截距这个参数）
  - 在HLM中自由度的定义：
    - HLM的回归方程不止一个，不仅有Level 1的一个方程，还有Level 2的一系列以Level 1的每个参数（包括截距和斜率）为因变量建立的方程，因此不同的回归系数因其对应的变量有不同的性质（或者说在HLM中所处的不同层级），其自由度也不尽相同。
    - 问题是：在使用R语言的lme4包和lmerTest包做HLM的时候，lmerTest输出的某个Level 2自变量的自由度（20多）远远小于它本身的Level 2样本量（400多），非常离谱，并且做的模型的自由度有可能都是小数而不是整数。不同软件在计算HLM自由度的时候，差异非常之大（vary enormously）。、
  - 目前计算HLM自由度的方法：

Conclusion: How to compute *df* in multilevel models

T-test should use the standard errors obtained from REML estimation, not from ML.

t = (estimate)/(standard error of estimate) = b/SE

For large sample size (i.e., number of groups; >= 30 groups):

Level-1 predictor (fixed slope; without cross-level interaction):

$df = N_{\text{sample size}} - p_{\text{level-1 predictors}} - q_{\text{level-2 predictors}} - 1$

$(= N_{\text{level-1 sample size}} - p_{\text{all parameters}})$

Level-1 predictor (random slope; with/without cross-level interaction), Cross-level interaction:

$df = K_{\text{groups}} - q_{\text{cross-level interactions}} - 1$

$(= K_{\text{level-2 sample size}} - q_{\text{all level-2 parameters for the specific level-1 predictor}})$

Level-2 predictor, Intercept:

$df = K_{\text{groups}} - q_{\text{level-2 predictors}} - 1$

$(= K_{\text{level-2 sample size}} - q_{\text{all level-2 parameters}})$

For small sample size (i.e., number of groups; < 30 groups):

Use the Satterthwaite approximation (Sattertwaithe, 1946; available in the R package *lmerTest*, or in SPSS MIXED):

df is estimated by the values of the residual variances

(correction for unequal variances and group sizes)

知乎 @包寒吴霜

Table 1. Computation of Degree of Freedom (*df*) in Multilevel Modeling

Predictor	Effect	*df*	Software
1. Large group sample size (*K* >= 30 groups)			
Level-1 predictor (γ_{10})	Fixed slope	$N-p-q-1$	HLM
	Random slope	$K-q_c-1$	
Cross-level interaction (γ_{11})	—		
Level-2 predictor (γ_{01})	—	$K-q-1$	
Intercept (γ_{00})	—		
2. Small group sample size (*K* < 30 groups)			
All predictors	—	Estimated by Satterthwaite's approximation	R (lmerTest), SPSS (MIXED), jamovi (GAMLj)

Note.
N = total number of observations (individual-level sample size)
K = total number of groups (group-level sample size)
p = total number of level-1 predictors in the level-1 sub-model ("Y_{ij} model")
q = total number of level-2 predictors in the level-2 sub-model for intercept ("β_{0j} model")
q_c = number of cross-level interactions in the level-2 sub-model for the *specific* level-1 variable ("β_{pj} model")

知乎 @包寒吴霜

Table 2. Common Examples

Level-1 Model	Level-2 Model	Type	Slope (X)	$df_{intercept}$	$df_{level\text{-}1\ (X)}$	$df_{level\text{-}2\ (W)}$
$Y_{ij}=\beta_{0j}+\varepsilon_{ij}$	$\beta_{0j}=\gamma_{00}+u_{0j}$	One-way ANOVA with random effects[1]	—	$K-1$	—	—
	$\beta_{0j}=\gamma_{00}+\gamma_{01}W_{1j}+u_{0j}$	Intercepts-as-outcomes model	—	$K-2$	—	$K-2$
$Y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+\varepsilon_{ij}$	$\beta_{0j}=\gamma_{00}+u_{0j}$ $\beta_{1j}=\gamma_{10}$	One-way ANCOVA with random effects	Fixed	$K-1$	$N-2$	—
	$\beta_{0j}=\gamma_{00}+u_{0j}$ $\beta_{1j}=\gamma_{10}+u_{1j}$	Random-coefficients regression model	Random	$K-1$	$K-1$	—
	$\beta_{0j}=\gamma_{00}+\gamma_{01}W_{1j}+u_{0j}$ $\beta_{1j}=\gamma_{10}$	Contextual model (fixed effects)	Fixed	$K-2$	$N-3$	$K-2$
	$\beta_{0j}=\gamma_{00}+\gamma_{01}W_{1j}+u_{0j}$ $\beta_{1j}=\gamma_{10}+u_{1j}$	Contextual model (random effects)	Random	$K-2$	$K-1$	$K-2$
	$\beta_{0j}=\gamma_{00}+\gamma_{01}W_{1j}+u_{0j}$ $\beta_{1j}=\gamma_{10}+\gamma_{11}W_{1j}+u_{1j}$	Full model[2]	Random	$K-2$	$K-2$	$K-2$

Note.
1. Also called: "null model", "intercept model", "variance component model".
2. Also called: "intercepts-and-slopes-as-outcomes model".

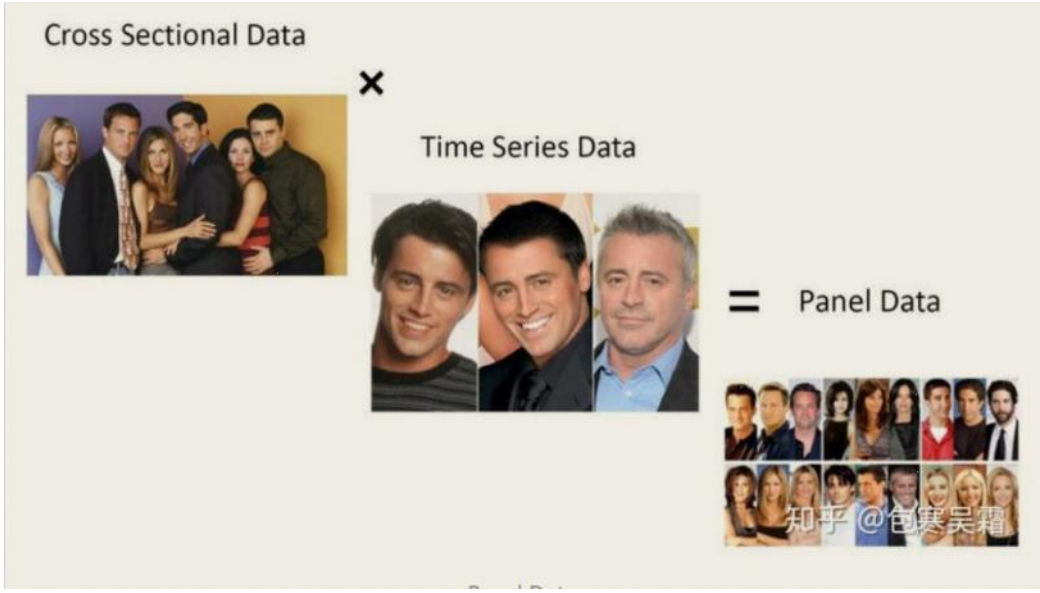
知乎 @包寒吴霜

1. HLM软件（软件本身叫做HLM）作为最经典的HLM统计工具，输出的都是由“简单计算法”估计的自由度，并且都是整数，不存在小数。这种方法理解起来也很简单，就是我们上面提到的原则：*df*（能够独立变化的数据量）= *N*（总数据量）- *k*（待估计的参数个数［包括截距］）。只不过N和k都需要具体看是哪一层的模型，对于Level 1的模型，与普通回归基本无异，而对于Level 2的模型，N不再是个体水平的总样本量，而是群体水平的“组数量”，k同样也变成了群体水平的参数个数（*注意：当level-1的某变量被设定为随机斜率，df计算略复杂，可参考上面的Table 1*）。由于HLM优于其他软件的一点在于它可以准确规定哪个变量属于哪个水平，所以虽然是简单计算法，但其实也是一种理论驱动的方法。
2. SPSS和R语言的lmerTest包均默认输出Satterthwaite（1946）的自由度近似估计值（approximation），这种算法是基于每一个变量在不同level的残差方差（residual variances）做的校正估计，严格来说是一种数据驱动的方法，主要用于校正方差不齐/异方差、每组内样本量不均衡、组数量较少等情况。这种近似估计往往得不到整数，而是会出现小数，但正常情况下，估计出来的自由度与简单计算法也是处于同一个数量级的，不会偏差太多。而当每组内样本量相等时，Satterthwaite的近似估计自由度与简单计算法得到的自由度就是一致的了，并且都是整数。
3. R语言的lmerTest包，除了Satterthwaite近似估计，还提供了Kenward-Roger近似估计，但后者不如前者好用（比如百万数量级的数据如果用KR估计就会直接崩溃）。
4. R语言的lme4包（最根源的HLM程序包，lmerTest只是调用了它并增添了假设检验结果）早期也会报告自由度和假设检验结果，但后来取消了这一功能，我推测可能是因为HLM自由度的估计尚存争议，所以希望用户自己来选择使用哪种自由度。
5. 自由度的估计方法，对显著性检验（p值）的影响其实并不大，但是对效应量计算（如t-to-r转换）的影响很大。因此用R作为统计工具（因为SPSS跑大数据模型实在太吃力了）更好，依然使用lme4和lmerTest包，但在自由度方面还是遵循HLM软件所使用的简单计算法，同时也看一眼lmerTest输出的

Satterthwaite校正结果，双重保障。

- 在ANOVA 方差分析中：
  - 基本特征：ANOVA既可以是GLM的特例，也可以是HLM的特例。在大多数情况下，被试间、被试内、混合设计ANOVA / ANCOVA都是GLM的特例——即自变量为分组变量的情况。总之，在ANOVA的框架里，所谓FE是指“固定截距”（GLM / OLS回归），而RE是指“随机截距”（HLM的特例）。
  - 例子1：例如，如果用OLS回归的方法分析一个单因素被试间设计ANOVA（假设该因素具有k = 3个水平），则等价于我们为这个因素生成了k – 1 = 2个虚拟变量（dummy variable；λ1 = [1, 0, 0]，λ2 = [0, 1, 0]；剩下的一个λ为冗余信息，舍去），然后使用这两个虚拟变量进行OLS回归，得到的结果与ANOVA一致。当然，除了dummy coding，还有其他几种不同的编码方式，如Helmert编码：λ1 = [2, -1, -1]，λ2 = [0, 1, -1]。但无论如何，我们为ANOVA里面的不同因素生成的编码变量本身只代表有限的分组信息，并不代表这些组别来自于一个更大范围的总体——比如最简单的“生理性别”，除了“男” “女”之外再无其他类别，故而用一个0/1虚拟变量就能表示，但这个变量也仅限于0和1，不可推广至2、3……因此，绝大多数的心理学实验设计都属于“固定因素模型”
  - 例子2：但还有一些情形，例如心理学研究者筛选实验材料，一般不可能穷尽所有可能性，而是按照一定的规则随机选取一些材料作为实验的stimuli，即“stimulus sampling”。这种情况就属于典型的“随机因素模型”——实质为HLM里面的“零模型”（具随机效应的ANOVA），并且同样要求材料的组别/水平数足够多才能进行有效的参数估计（即保证足够大的Level 2自由度df，从而保证power）。
  - 例子3：对于被试内/重复测量设计的ANOVA，同样可以考虑使用HLM（或者被认知心理学背景的学者称为“线性混合模型 LMM”，本质都一样）。此时，被试个体将作为Level 2的clustering variable，并且是随机截距（1 | subject），而重复测量的条件将作为Level 1自变量，视情况设为固定或随机斜率。使用HLM处理重复测量实验数据具有一定的优势，尤其体现在重复测量条件很多、很复杂、可以为连续变量的情况，例如眼动数据、脑电数据等等。

- 在Panel Data Model 面板数据模型中
  - Panel data的数据结构是N个“个体”（广义的个体，可以是人，也可以是宏观层面的省份、城市）× T个时间点（小到day-level，大到year-level）。这种N × T的四四方方的面板/追踪数据其实是综合了横截面（N）和时间序列（T）两个维度，可以解决单独的横截面或时间序列数据所不能解决的问题，并且在计量经济学里面特别常用。更重要的是，究竟该使用FE还是RE，也是一个在处理panel model时需要考虑的基本问题。

- 

  - PLM所谓的固定效应FE，在本文讨论的所有统计模型中是唯一的一个有点“名不副实”的术语——因为即使在FE的PLM里面，个体效应也依然是「异质」而非真正「固定为同一个值」的（陈强, 2014《高级计量经济学及Stata应用》）。换句话说，在PLM里面，无论FE还是RE，个体截距（相当于HLM里面Level 2不同组各自的截距）都是“异质性截距”（ [公式] ）。而区别在于，FE和RE对于这种“异质性截距”有着不同的假定，相应的也有不同的模型估计方法：
  - FE（异质性[非随机]截距）：由于PLM数据有个体和时间两个维度，所以FE也分为个体固定、时间固定、双固定。例如个体固定，可以类比于ANOVA里面把不同组别用虚拟变量来表示。我们可以使用“最小二乘虚拟变量回归法”（Least Square Dummy Variable, LSDV）来分析PLM，那么在模型估计的时候，LSDV的做法就是给N个个体生成N – 1个0/1虚拟变量，然后将这些虚拟变量与主要的预测变量一起纳入回归方程，做OLS回归（并且常常需要计算cluster稳健标准误）。时间固定与双固定的做法与之类似。除了LSDV，还可以使用“均值离差法”、“一阶差分法”等分析PLM，其中，均值离差法较为常用，其做法是计算出每个个体在T个时间点上的“个体内均值”，然后用原始观测值减去个体内均值进行均值离差校正（相当于做了一个跨时间的组中心化处理；time-demeaning），最后以自变量和因变量的离差值进行回归，分析得到的结果与LSDV基本是一致的。因此总体来看，所谓固定效应FE，在PLM的框架下可以理解为做的是虚拟变量OLS回归（LSDV）。由于在多数的经济数据中，个体不可观测的异质性截距（[公式]）往往与解释变量（[公式]）有关或相互干扰，FE的这种虚拟变量回归的做法可以很好地控制并排除那些不可观测的个体差异的影响，从而可以在一定程度上解决遗漏变量的问题，提高模型的准确性。
  - RE（异质性[随机]截距）：上面所说的FE允许个体异质性截距 [公式] 与解释变量 [公式] 存在相关；相比之下，RE有着更严苛、也更难满足的假设，即假设 [公式] 与所有自变量 [公式] 均不相关，并且 [公式]（残差）是一个服从正态分布的随机变量。RE模型也有相应的估计方法，例如可行广义最小二乘法（FGLS）、最大似然法（ML）等。如果使用最大似然法，其结果将与HLM随机截距模型的结果一致。
  - 虽然一些统计检验（如Hausman检验）可以在一定程度上告诉我们应该使用FE还是RE，但是，至少在经济学领域，研究者通常都会使用固定效应FE。为什么呢？首先，这是因为无论Hausman检验的 [公式] 是否成立，FE得到的参数估计量都是“一致的”（i.e., 点估计收敛于真实的参数值），只不过如果 [公式] 成立，RE会比FE更“有效”（i.e., 回归系数的标准误更小，也就是“更显著”），但假如存在异方差，RE也并非最有效。其次，经济学往往关注国家、省份、城市等宏观层面的规律。显然，省份这种分组变量本身就可以代表其总体（即一个国家的省份是固定的），而不是一种随机抽样；同时，省份在一个变量上的“基线水平” [公式] 往往也会与其他变量 [公式] 存在相关。因此综合来看，经济学的宏观数据天然更适合用FE（异质性[非随机]截距）。

• 在Meta-Analysis元分析中

- 一般情况下，元分析的基本分析单位是从已有文献中提取出来的效应量，包括但不限于原始的均值M、Cohen's d、相关系数 r、odds ratio（统计量–效应量的相互转换 | 元分析基础）（https://zhuanlan.zhihu.com/p/47849067）。例如相关系数 r，我们在做元分析的时候不仅需要输入 r 本身，还需要输入研究的样本量 n。这是因为，无论我们选取哪一种效应量做元分析，都需要知道这个效应量（effect size）在最初由原始研究得出的时候具有多大的取样误差（sampling variance; within-study error）。样本量越大，取样误差越小，得到的效应量置信区间（95% CI）也就越精确。
- 换一种视角来看，已有文献中的效应量，本身可视为“回归斜率”（回忆：众多统计方法的本质其实都是回归分析，即使是组间差异也可以转换为回归斜率或r-family的效应量）。更进一步来说，这种原始的回归斜率或者说效应量，放在HLM的框架里面就变成了Level 1的斜率；同时，取样误差作为效应量标准误的度量，也就变成了Level 1的残差。假设我们拥有这些研究在个体层面的原始数据，那么毫无疑问，我们会做HLM：将不

同的研究作为Level 2的分组变量，而被试个体作为Level 1。

- 然而，骨感的现实是，我们无法获取原始数据（大概因为我们比较穷也比较懒）。所以我们会去做“元分析”（可以戏称为“穷人/懒人的HLM”），也就是从已有文献里面抠出来一个个效应量，看看它们之间是否一致、平均效应是否存在（这还真不是一个懒人能完成的任务）。于是，我们建立的元分析模型，本质上就是一个只有Level 2的HLM。但由于我们还获取了效应量对应的取样误差，所以我们做的并不是一个纯粹在组层面的GLM回归，而确确实实是一个HLM——只不过Level 1已经由算好了的效应量及其误差代替了而已。
- 因此，元分析模型里面的“截距”（如果你做过的话就会看到），本质上并不是“截距”，而应该理解为“平均的Level 1的斜率”（平均效应量）。于是，元分析的话语体系里所谓的“固定效应模型”本质就是HLM的“固定斜率”（即假设效应量之间不存在异质性，是同质的），而“随机效应模型”则是HLM的“随机斜率”（即假设效应量之间存在异质性，依研究而变）。当然无论是FE还是RE，最终我们都会得到一个平均的、汇总的效应量的点估计及其区间估计。
- 总体来看，随机效应RE的元分析模型更受推崇——不仅因为RE更符合实际情况（不同研究的异质性）、在估计时更加保守、更不易被极端大/极端小样本量的研究“带跑偏”（平均效应的估计更均衡、更稳健），而且如果我们进一步纳入Study-level的解释变量，还可以做“元回归”（meta-regression）来解释为什么效应量在不同研究之间会存在差异。例如，是不是研究取样的性别比例、平均受教育水平等因素解释了效应量的差异？元回归理解起来其实也很简单，相当于在HLM随机斜率的基础上检验跨层交互作用——Study-level的变量在元分析的模型里面是predictor，其本质则是HLM Level 2的moderator，调节的是Level 1的斜率（效应量）。
- 总之，除了一些少数情况（包括：①我们有很强的假设认为纳入元分析的不同研究之间是同质的，②我们并不想把结论推广到更大的总体），大多数情况下做元分析还是建议使用随机效应RE。而且，研究者并不建议根据“效应量异质性检验”的结果（如Q、I^2等统计量）来决定采用FE还是RE——不要数据驱动，要理论驱动！顺带一提，jamovi作为一款新兴的统计神器，也能直接做元分析，非常方便（我是见人就安利jamovi的，哈哈哈）。