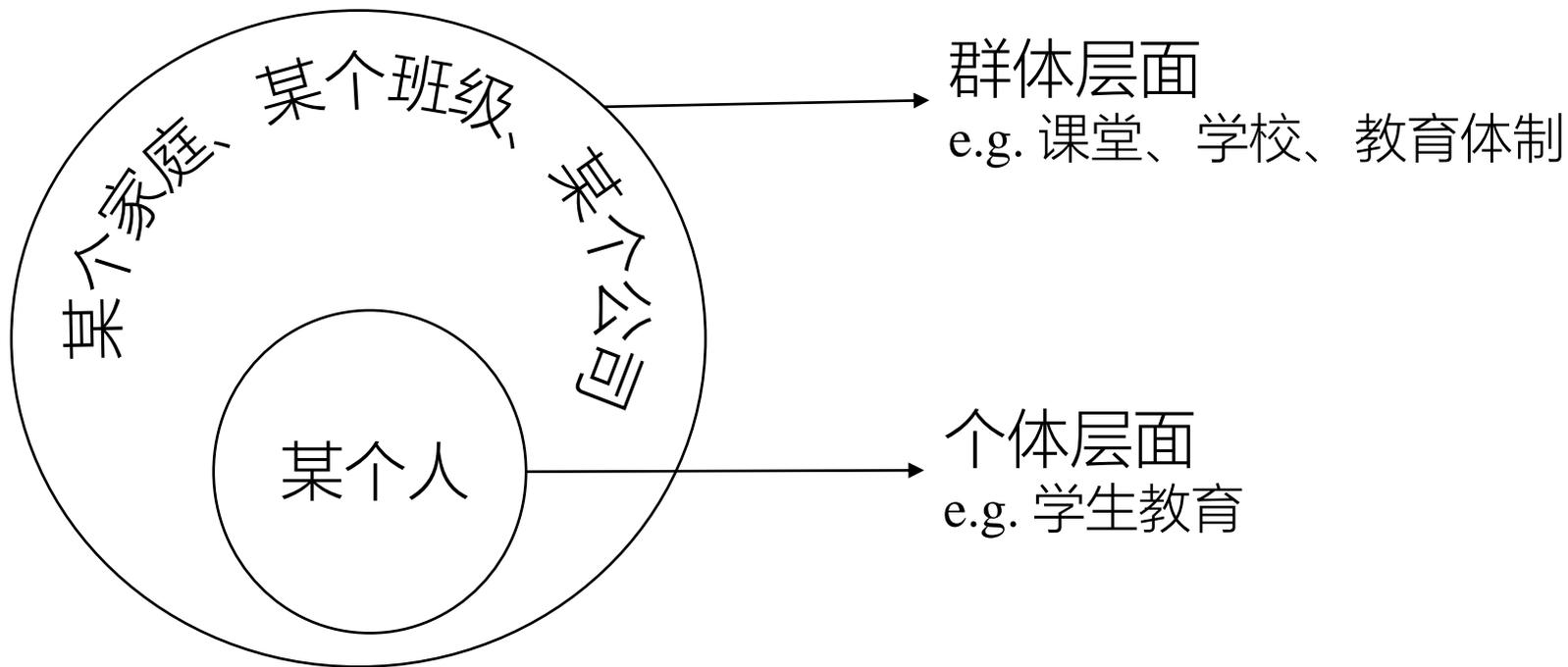


# 多层次的数据结构

## 多层次模型 ——看一场3D电影

报告人：黄颖诗

# 多层次模型



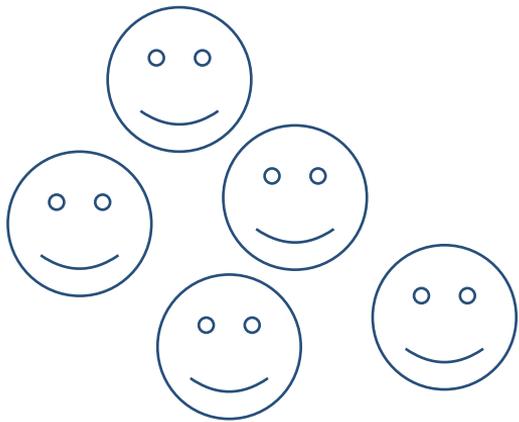
# 多层次模型

- 生态谬误 (ecological fallacy)
  - 群体间的关系被假设同样存在于个体之间 (Freedman, 1999)。  
比如, 教育和经济水平越发达的地区生育水平越低, 并不一定能够引申为个人受教育水平越高、收入越多, 生育数量越低。
- 原子谬误 (atomistic fallacy)
  - 群体信息被不正确地从个体信息中推论出来 (Hox, 2002)。比如, 流动人口的犯罪率高不等于流动人口比例高的城市犯罪率高。

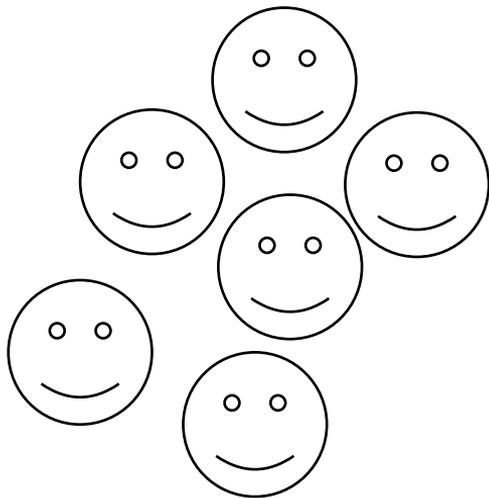
# 多层次模型

- 相同背景下的个体，彼此之间存在某些相似性。  
→ 误差项必然相关

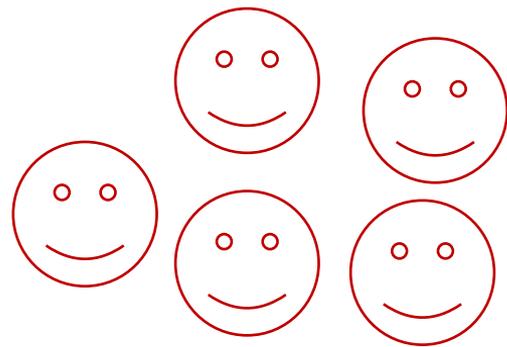
班级一



班级二



班级三



# 多层次模型

- 基于一系列非同一层次的自变量对因变量的值进行估计。如，学习时间（个体特征）对阅读能力的影响，还应考虑班级学习氛围（群体特征）。

第一层： $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$  (如：X = 学习时间)

典型的OLS回归

第二层： $\beta_{0j} = r_{00} + r_{01}Z_j + \mu_{0j}$  (如：Z = 班级学习氛围)

将截距和斜率  
作为第二层自  
变量的结果

$$\beta_{1j} = r_{10} + r_{11}Z_j + \mu_{1j}$$

# 多层次模型

第一层:  $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$

第二层:  $\beta_{0j} = r_{00} + r_{01}Z_j + \mu_{0j}$

$\beta_{1j} = r_{10} + r_{11}Z_j + \mu_{1j}$

$Y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}] + [\mu_{0j} + \mu_{1j}X_{ij} + \varepsilon_{ij}]$

固定效应部分

随机效应部分

## 多层次模型

- 现在有几层数据？其中有几层被模型化？
- 各层次分别有多少个自变量需要考虑？
- 将一层的斜率还是截距，或是两者共同作为二层特征的结果？
- 哪部分引入随机效应？随机截距还是随机斜率？

# 多层次模型

- 什么情况下一定要用多层次分析方法？

- 理论方面

所使用的研究框架或建构的理论为多层次运作。

- 统计方面

数据的结构特征，出现嵌套结构意味着各项观察值并不独立，即数据的聚类本质。

- 实证方面

画图技术、组内相关系数（ICC）。

# 多层次模型

- 什么情况下一定要用多层次分析

- 理论方面

所使用的研究框架或建构的理论为：

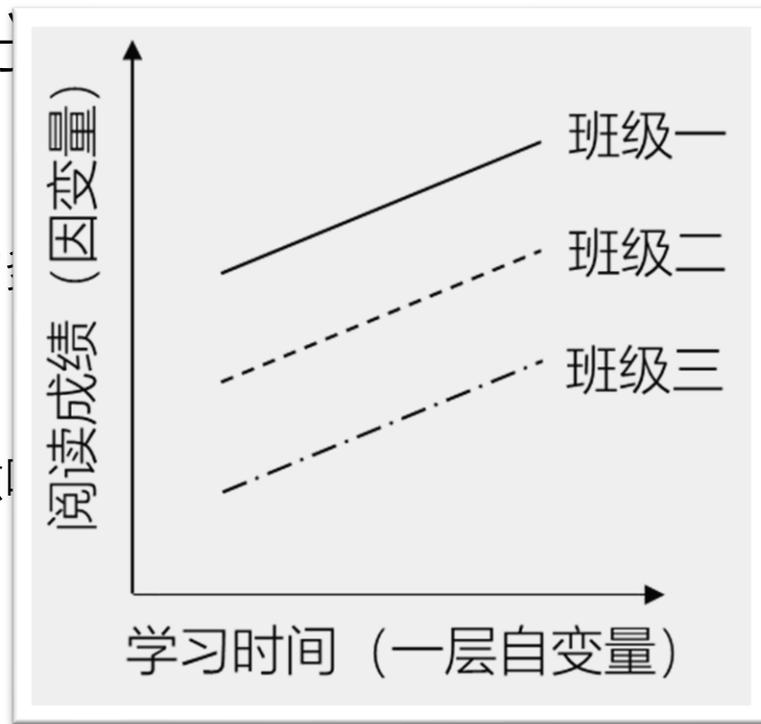
- 统计方面

数据的结构特征，出现嵌套结构意

即数据的聚类本质。

- 实证方面

画图技术、组内相关系数（ICC）。

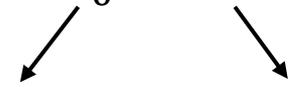


# 多层次模型

- 组内相关系数 (ICC)
  - 在没有任何解释变量的情况下，估计组间方差占总方差的比例。

$$\begin{aligned} \text{零模型: } Y_{ij} &= \beta_{0j} + e_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j} \end{aligned}$$

$$ICC = \rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_{\varepsilon}^2}$$



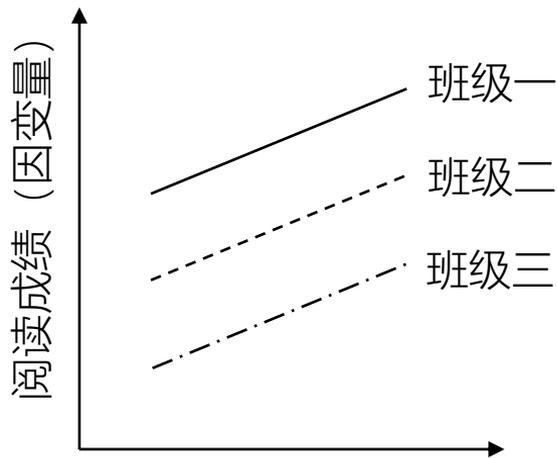
第二层的方差 (组间)      第一层的方差 (组内)

小于0.059，相当小的组内相关系数  
0.059-0.138，中度相关  
大于0.138，高度的组内相关

# 多层次模型

- 随机截距与随机斜率

截距变动

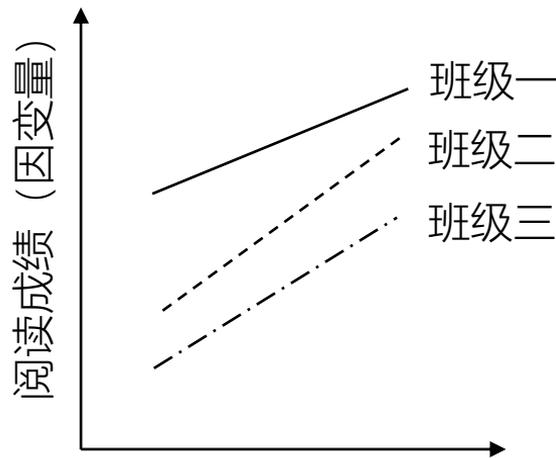


学习时间 (一层自变量)

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = r_{00} + r_{01}Z_j + \mu_{0j}$$

$$\beta_{1j} = r_{10} + r_{11}Z_j$$



学习时间 (一层自变量)

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = r_{00} + r_{01}Z_j + \mu_{0j}$$

$$\beta_{1j} = r_{10} + r_{11}Z_j + \mu_{1j}$$

截距和斜率变动

# 多层次模型

- 适用于多层数据的模型
  - 发展模型——不同时间点的数据嵌套于个体研究对象
  - 随机效应模型——把固定效应推广到随机效应
  - 广义分层线性模型——对因变量进行转换并引入合适的误差分布

# 多层次模型

- 分析软件
  - HLM ( Raudenbush et al., 2000 )
  - Mlwin ( Rasbash et al., 2000 )
  - VARCL ( Longford, 1988 )
- 当代发展
  - 解决嵌套数据的问题
  - 因变量从单一变量到多变量模式
  - 潜在变量模型

## 多层次的数据结构

Thank you for your attention!

*2019-03-05*