病例全在第一批，对照全在第二批，这份数据还能解释哪些差异?

用三张小型样本表比较完全混杂、单个跨批次样本和两批均衡设计，亲手计算设计矩阵、条件对比的可估性和删供者敏感性。



运行教学示例

Python 3.10 或以上，仅需 numpy。解压后运行 python -m pip install -r requirements.txt。

```
python run.py --input example --out results
```

在解压后的资料包根目录运行。example 是教学输入，results 保存本次输出，expected 保存可供核对的参考结果。

输入字段

samples.tsv 三种教学设计。每行是一份捕获，包含供者、病例对照、技术批次及示例样本表达。

核对输出，再换自己的数据

design_summary.tsv、design_columns.tsv 矩阵秩、系数列数、条件对比的可估性，及逐行编码。示例表达只演示可估差值。

condition_by_batch.tsv、within_batch.tsv 各条件与各批次的捕获数和独立供者数；批次内实际可比的病例对照差。

technical_links.tsv、leave_one_donor.tsv 跨批技术重复的供者及逐位移除后的条件对比状态。

用示例验算

1 六位供者完全混杂时矩阵秩为 2、列数为 3，校正批次后的条件差记为 NA。

2 单一桥接时有 7 份捕获、6 位独立供者，可估的教学差为 1，去掉 P1 再度不可估。

3 均衡设计两批各有 2 病例、2 对照，每批教学差均为 2，加性模型给出同样的 2。

换用真实数据

以独立样本为单位导出条件、供者、捕获与批次表。遇到纵向或配对设计另设模型；样本级实际表达需按相应计数和重复结构建模。

例中表达仅用于演示加性均值差，不输出 P 值。跨批同一供者的两次测量不能算作两名独立供者。

方法资料

DESeq2 的设计矩阵与完全混杂示例

R 的模型矩阵编码说明