

《传播统计学》

抽样

教师：林志良

邮箱：linzhl@nfu.edu.cn

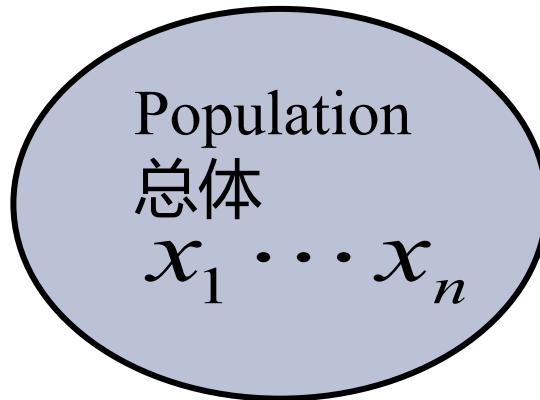
个人网站：www.zhilianglin.com

目录

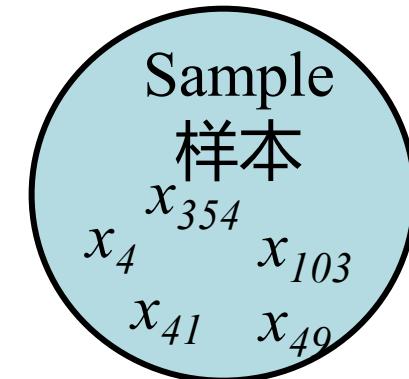
- 总体与样本：相关概念、生活实例
- 抽样：误差、抽样方法、软件实操（Excel + SPSS）

- **总体** (population) : 需要研究的对象的全体。
- **个体** (unit) :每一个具体研究对象。
- **样本** (sample) :从总体中抽出的用以推测总体的部分对象的集合。
- 样本中包含的个体数，称为**样本的容量** n 。一般把容量 $n \geq 30$ 的样本称为**大样本**；而 $n < 30$ 的样本称为**小样本**。

抽样 (sampling)



我们从总体中
「选出」一批人
来观察



从观察样本的结
果，来推论总体
的性质

推断统计是用样本数据做出关于总体的一般性陈述的方法。



总体与样本

相关概念

- | | | |
|------------|----|-------------|
| ➤ 总体 | —— | 样本 |
| ➤ 参数值 | —— | 统计量 |
| (parament) | | (statistic) |

固定数值

根据样本的
变动而变动



希腊字符

英文字母

总体参数与样本统计量的统计符号对比

Measure 测量	Population Parameter 总体参数	Sample Statistic 样本统计量
Mean 平均值	μ	\bar{X}
Variance 方差	σ^2	S^2
Standard Deviation 标准差	σ	S

街头拦行人进行调查/记者上街“随机采访”，样本有何问题？

- 出现在商场（街上）的行人并不能代表总体的基本状况。
- 我们对于受访者是有挑选的！

课堂上的点名是随机的吗？

- 名字简单容易被点。

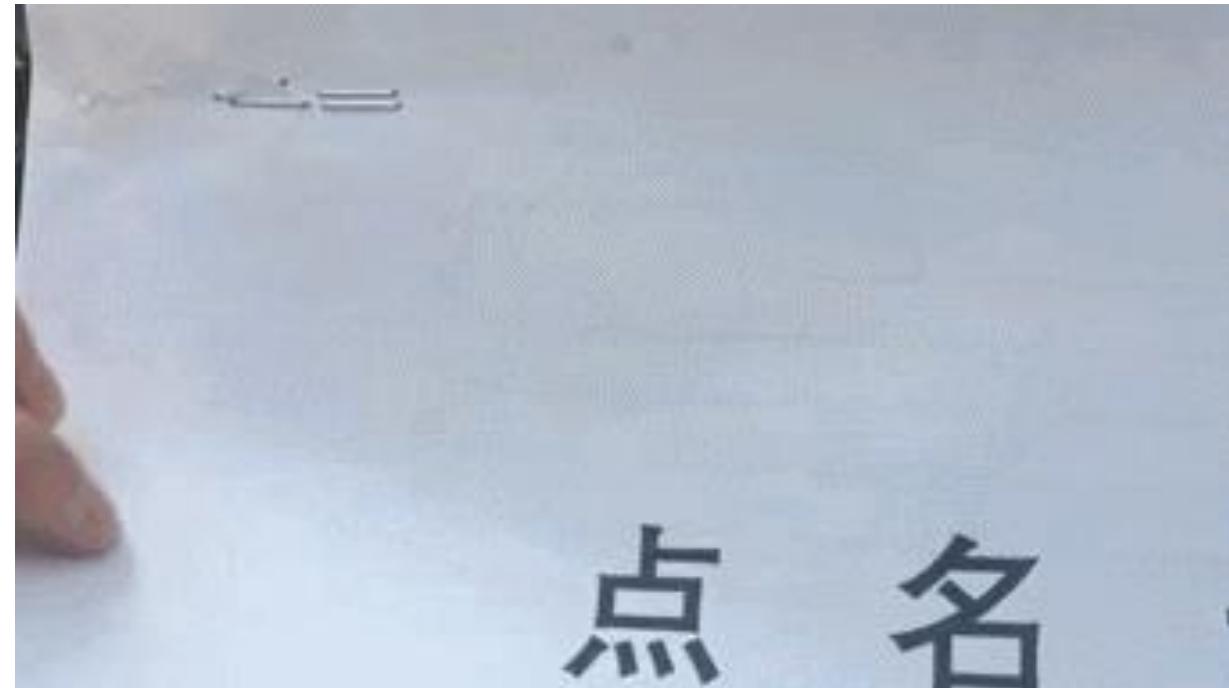
序号	姓名	性别		
			1	2
16	欧阳锋	男		
17	万俟高	男		
18	韦小宝	男		

中间那个名字，看着就不想点



课堂上的点名是随机的吗？

- 位置分布居中容易被点。



课堂上的点名是随机的吗？

- 视觉重点容易被点。

23	杰克隽逸	女
24	古力娜扎	女
25	杨紫	女
26	欧阳娜娜	女

三长一短选最短

27	鹿晗	男
28	吴磊	男
29	迪丽热巴·迪力木拉提	女
30	王源	男

三短一长选最长

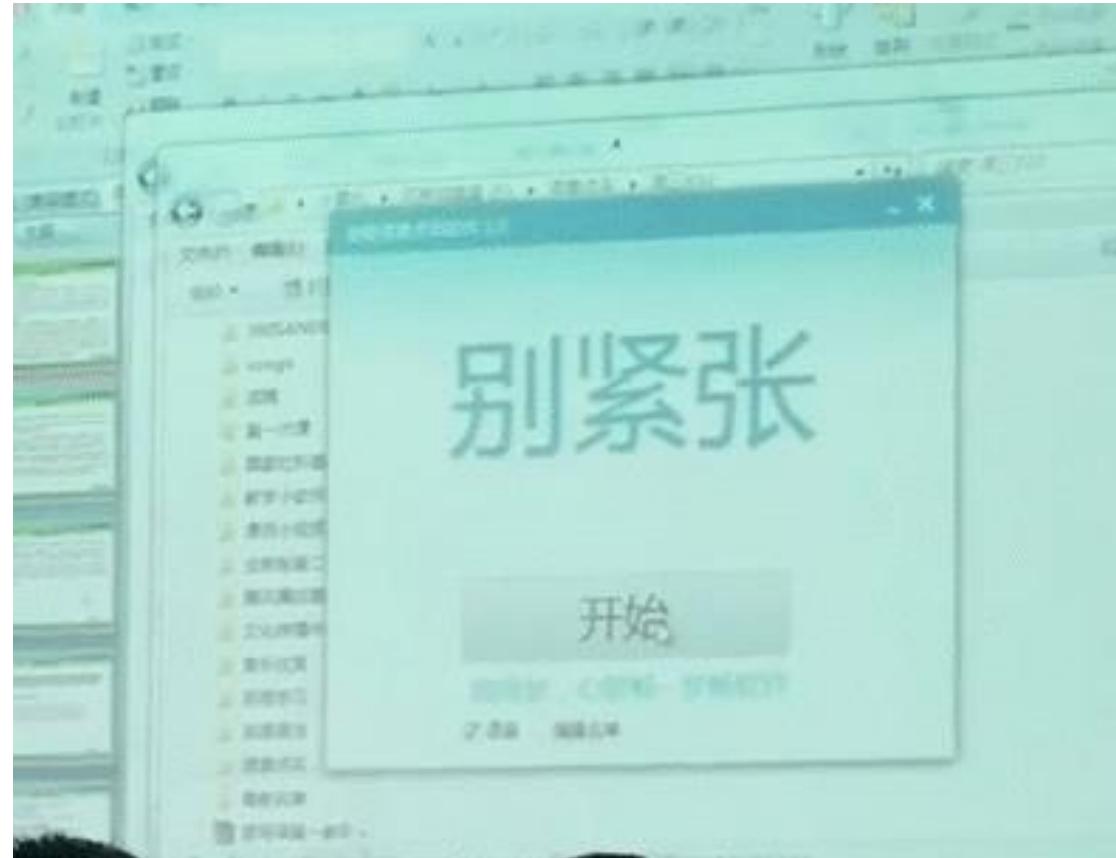
课堂上的点名是随机的吗？

学号点名



课堂上的点名是随机的吗？

科技点名



误差

调查研究中的误差

- **抽样误差** (sampling error)
- **非抽样误差** (non-sampling error)

误差

抽样误差

- **抽样误差** (sampling error) : 是抽样所造成的误差，抽样误差使得样本结果和普查结果不同。
 - **随机抽样误差** (random sampling error) : 样本统计量和总体参数之间的差距，是在选取样本时由随机性造成的。
 - **涵盖不全**。
 - **糟糕的抽样方法**，比如**自愿回应**。



虽然退出现代社会有时会让人觉得无聊，不过只要能加剧民意调查的涵盖不全问题，泰德就觉得这样做是值得的。

案例：糟糕的样本

- 美国某地，多年来只有一家公司获准提供急救服务。1999年，当地报纸《小镇消息》请人家致电谈一谈他们**是否同意让那家公司继续垄断经营**。这种致电民意调查一般是**自愿的**：打其中一个电话号码是投赞成票，而打另一个电话号码是投反对票。
- 结果该报接到了3763通电话，表明这个回题得到了非同寻常的关注。调查表明，其中有638通电话是从该公司办公室或其管理人员的家中打来的。

案例：糟糕的样本

- 该报的民意调查依赖 “**自愿回应**” (voluntary response) , 即人们主动打电话过来, 而不是用报纸来抽取样本, 结果就是 “**有偏的**” (biased) 。样本中赞成继续垄断的人的比例提高。自愿回应法吸引到的样本是对该话题有强烈感受的人。

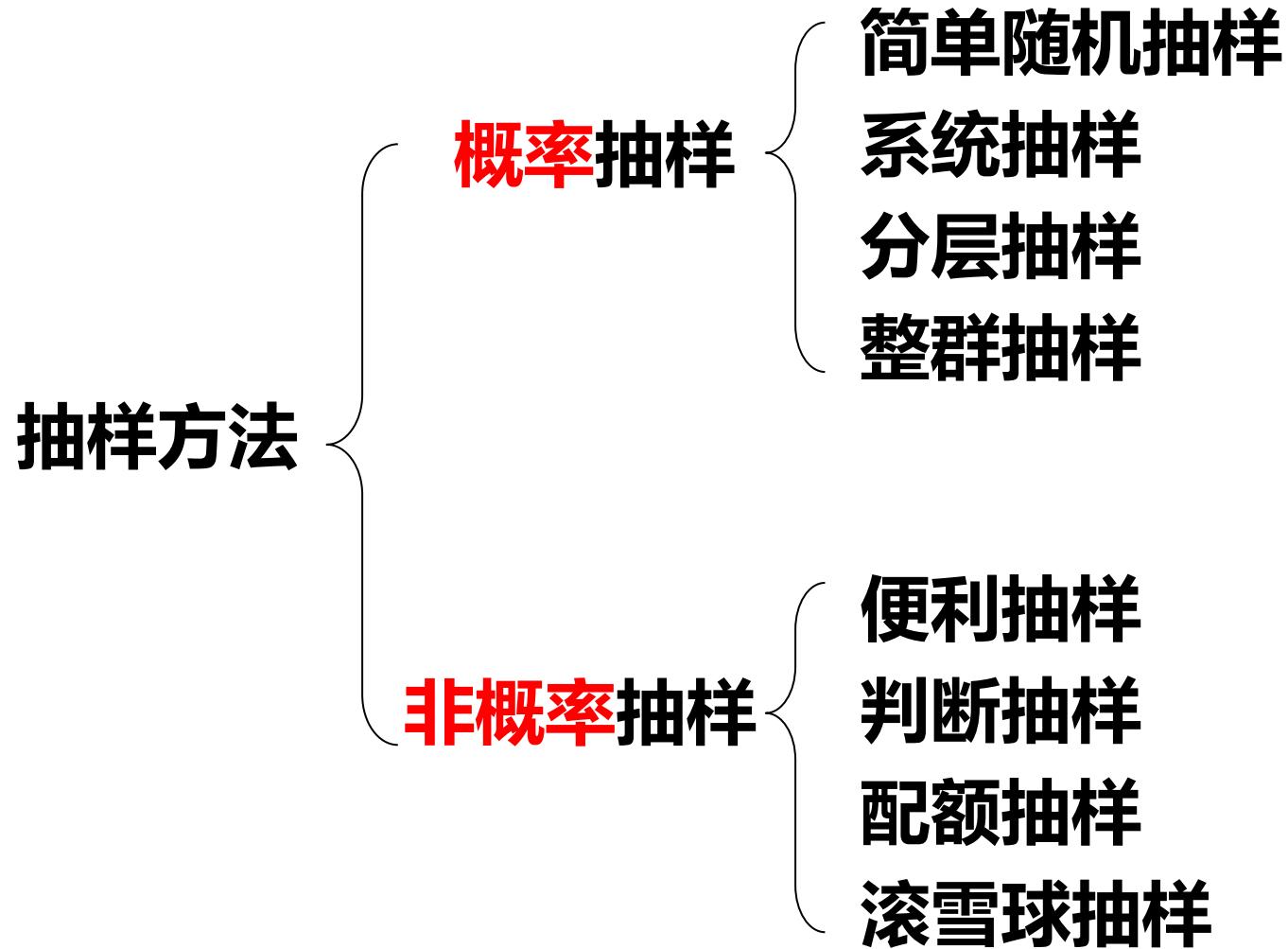
- **非抽样误差** (non-sampling error) : 和“从总体中抽取样本”这个做法无关的误差，非抽样误差即使在人口普查中也有可能出现。
 - **处理误差** (processing error) : 机械化工作时犯的错误。例如，计算错误或将受访者的回答输入电脑时犯的错误。
 - **回应误差** (response error) : 受访者谎报或因记忆错误导致回答错误。

案例：“布拉德利效应”

- 来自洛杉矶前黑人市长汤姆·布拉德利，即使在选举前最后一天的民意调查中领先，他仍在1982年的加州州长选举中落败了。
- 调查机构相当肯定，一部分选民在接受访谈时说了谎，因为他们不愿承认自己反对黑人候选人。这个现象被称作“**社会合意性偏误**”（social desirability bias）或“**布拉德利效应**”。



抽样方法



概率抽样

抽样框

- 一个抽样框是构成总体的全部单元的“家庭聚会”，样本可由此抽选出来。
- 不借助抽样框，研究者就难以保证以同等概率抽取样本。当样本框不能完全代表所要研究的人群时，则这二者之间的差异(抽样框误差)就会影响分析结果的可靠性。这种误差是构成抽样误差的主要来源之一。

概率抽样

简单随机抽样

- 大小为 n 的“简单随机样本”(simple random sample, SRS)是指有 n 个个体的样本。其样本选取的原则是，**总体中任意一个个体的中选概率相同。**
- **两个步骤选取简单随机样本**
 - 第一步：标签为总体中的每个个体分配一个数字标签。
 - 第二步：软件或表使用电脑软件或随机数字表，随机选择样本。

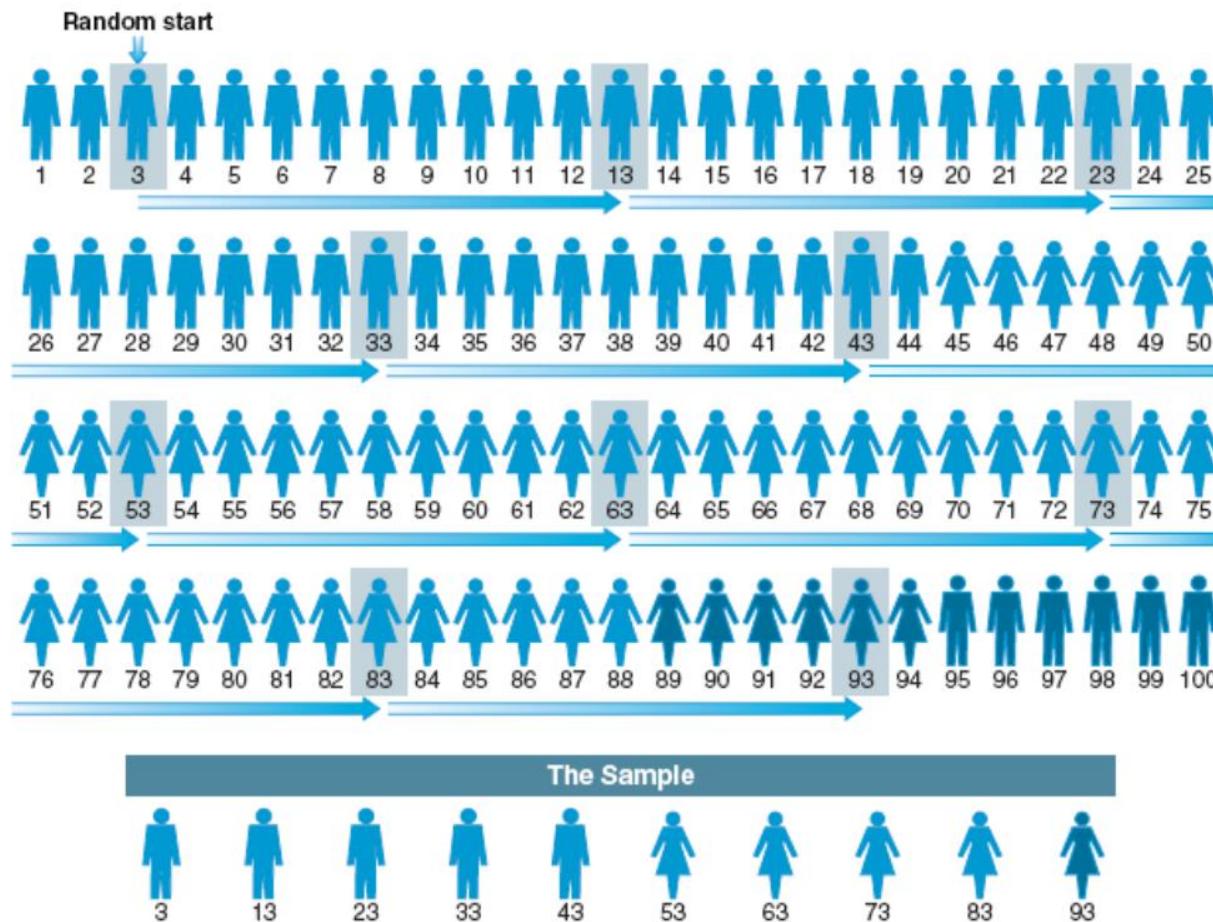
概率抽样

系统抽样 (systematic sampling)

- 系统地选择完整名单中的每第K个个体组成样本。
- 例如，总体为1000个个体，我们需要100个个体，就可以选择每第10个个体作为样本。
 - 随机给1000个个体编上1-1000的序号。
 - 抽取编号为3, 13, 23...的个体

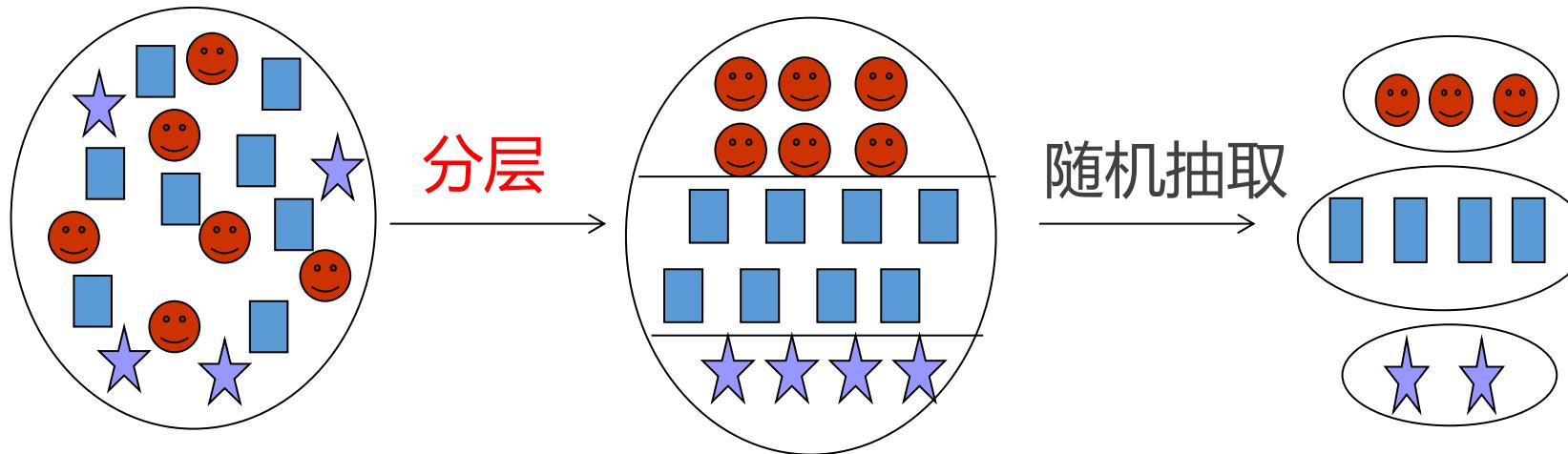
概率抽样

系统抽样 (systematic sampling)



概率抽样

分层抽样(stratified sampling)



概率抽样

我的行动能力



我的知识储备



我的脾气



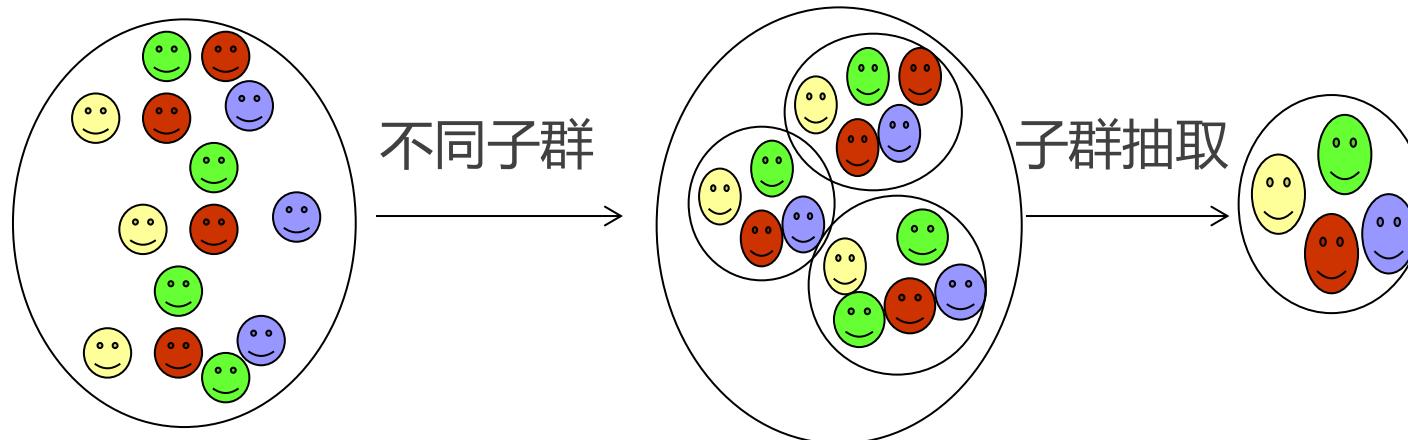
概率抽样

分层抽样的基本步骤

- **第一步：**将抽样框架中的个体先**分成若干群**，被称为层分层的标准是，你对于这些层有特别的兴趣，或者同一层中的个体有相似的性质。
 - 如：依照**科系**对教师进行分类；依照**班级**对中学生进行分类； ...
- **第二步：**每层各取一个**简单随机样本**，把它们合起来就是我们要的样本。

概率抽样

整群抽样 (cluster sampling)



概率抽样

整群抽样

- 它是从总体中随机抽取一些小的群体，然后由抽出的若干个小群内的所有元素构成调查的样本的方法。
- 假设某大学共有100个班级，每班都是30名学生，总共有3000名学生。现要抽300名学生作为样本。如果我们采用整群抽样的方法，就不是直接去抽一个个的学生，而是从全校100个班级中，采取简单随机抽样的方法（或是系统抽样、分层抽样的方法）抽取10个班级，然后由这10个班级的全部学生（300名）构成样本。

非概率抽样

便利抽样 (convenience sampling)

- 从总体中抽样时，如果抽取最容易得到的个体，这种方法就叫作“**便利抽样**”。
- **典型例子**：街头拦截路人进行调查/采访；教师把上课学生作为调查对象。

非概率抽样

便利抽样

- 我们要谨慎通过便利样本去**推论总体**，但用便利样本分析**变量之间关系**在学术界接受度其实并不低。
- **适用场景**：调查对象流动性强（如黄金周的旅客）；正式调查前的**预调查**。

非概率抽样

配额抽样 (quota sampling)

- 按照不同属性的个体在总体中的比例来分配样本数量。
- 例如，如果要通过配额抽样的办法调查南方学院在校生，需要知道南方学院男女比例、各学院人数比例、生源地比例等数据，然后在数据收集阶段让样本的人口特征比例和总体比较接近。

非概率抽样

目的性抽样 (purposive sampling)

- 又称**判断性抽样** (judgmental sampling) , 根据对总体如对总体个体以及研究目标的认识, 来选择合适的样本。
- 例子: 大学教务处教学检查小组认为某学院教学质量较差, 因此在教学质量抽查时着重挑选该学院。

非概率抽样

滚雪球抽样 (snowball sampling)

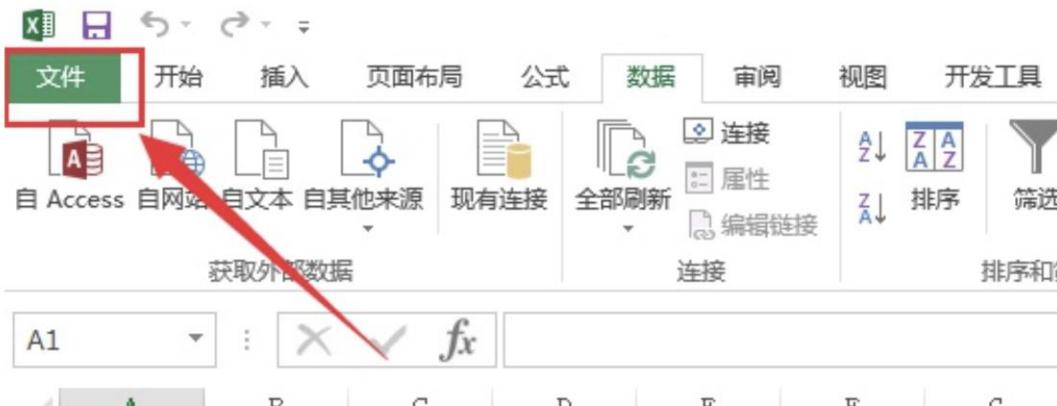
- 先搜集目标群体少数成员的数据，然后再向这些成员询问有关信息，找出他们认识的其他成员。
- 比较适合于对边缘群体进行调查。

Excel抽取随机样本

准备工作

Excel数据工具包准备

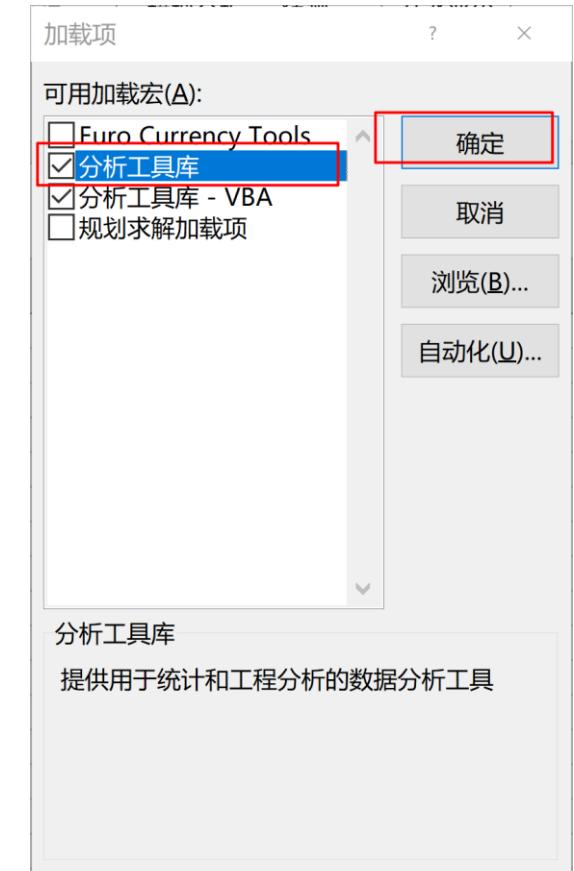
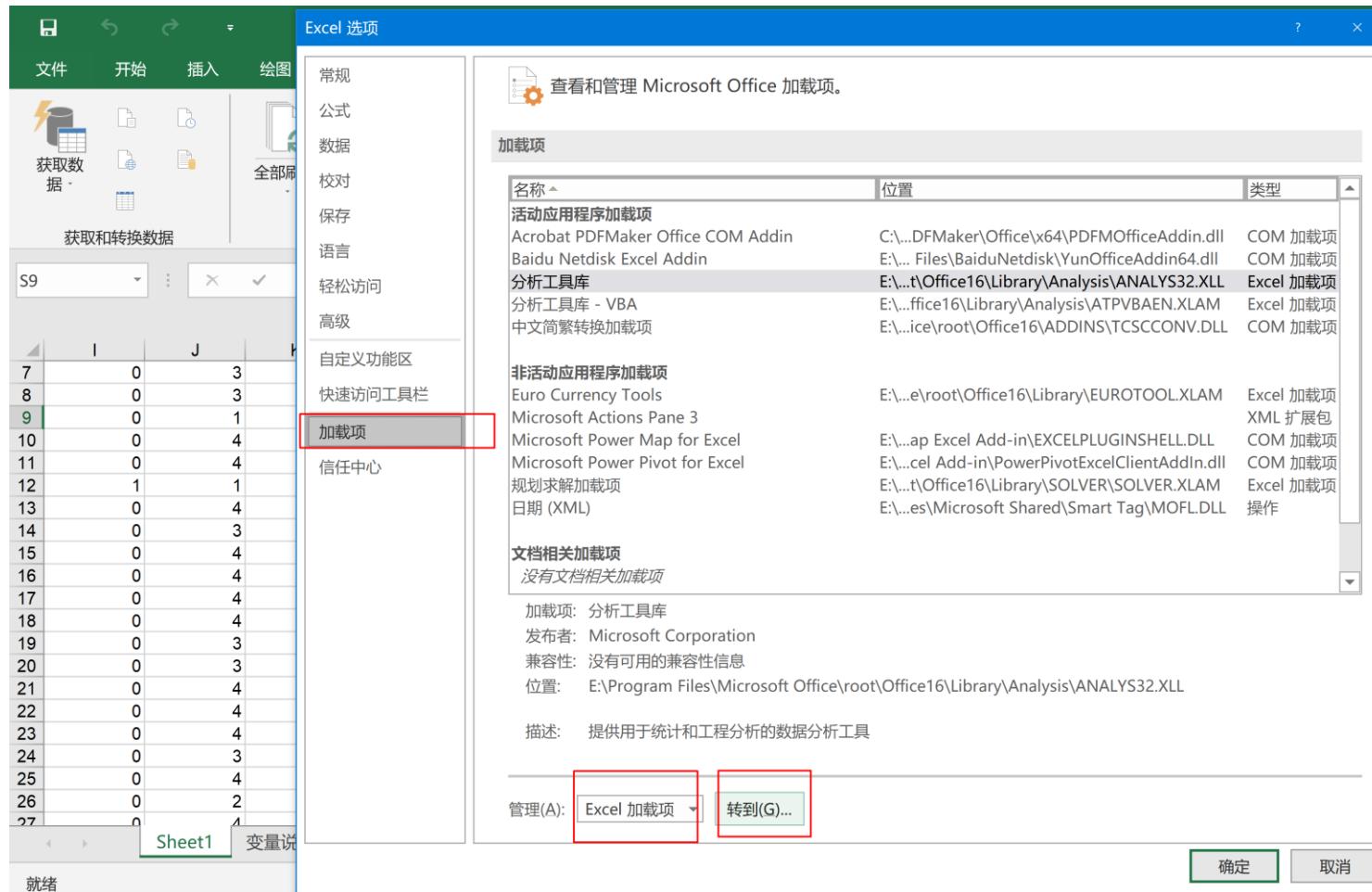
【文件】>【选项】>【加载项】>【Excel 加载项】>【转到】>勾选【分析工具库】>确定



Excel抽取随机样本

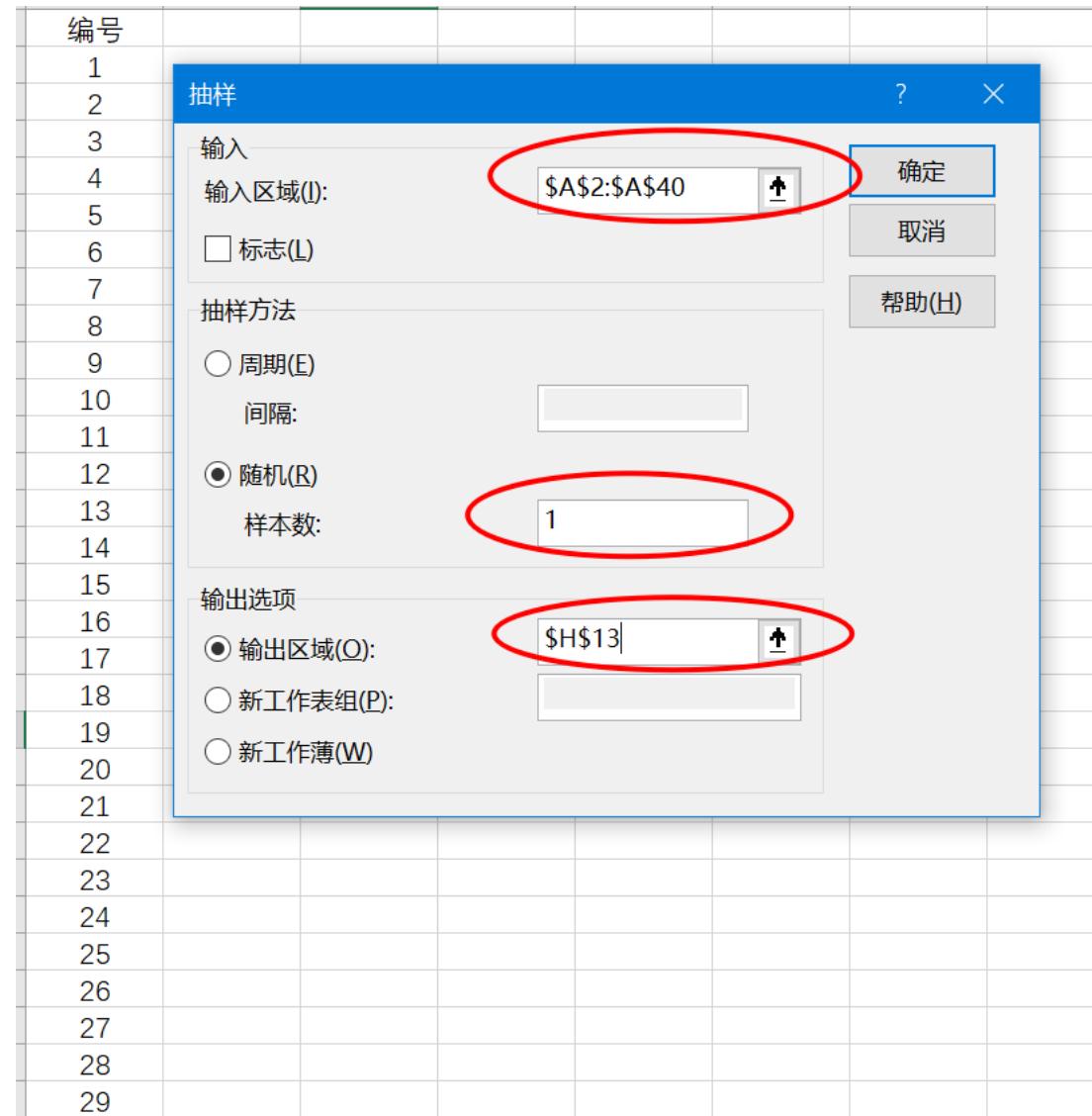
准备工作

Excel数据工具包准备



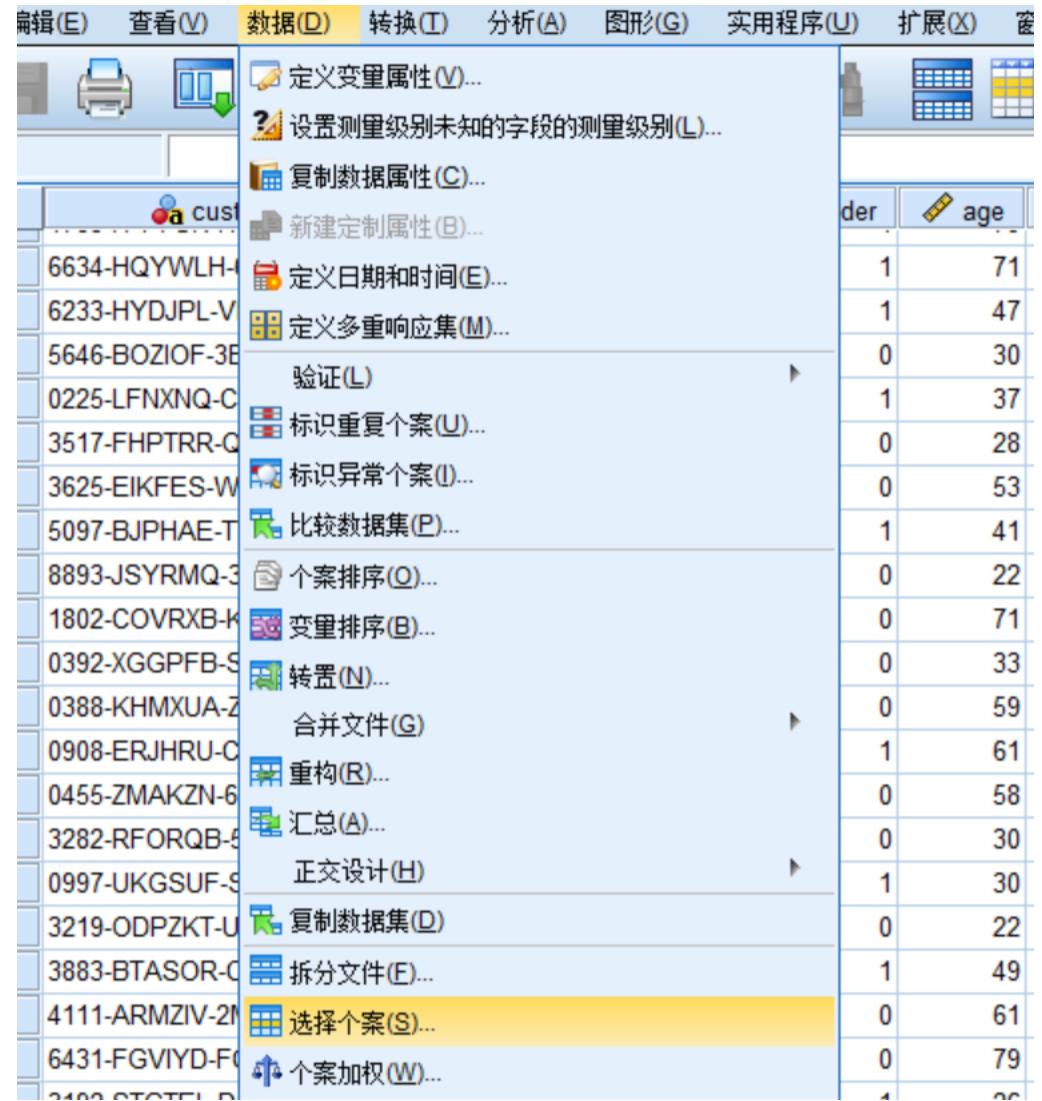
Excel抽取随机样本

Excel抽取随机样本



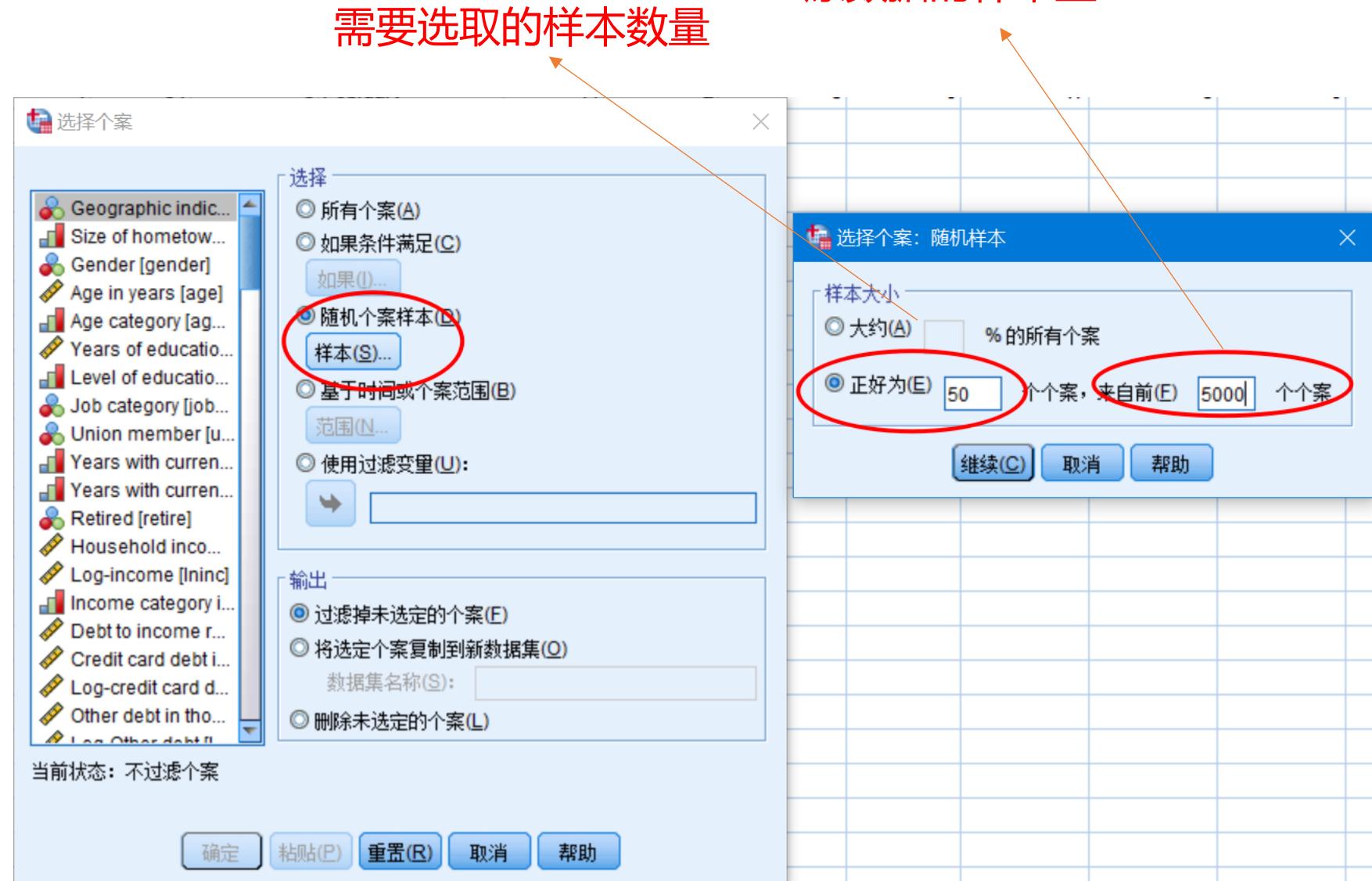
SPSS抽取随机样本

【数据】>【选择个案】



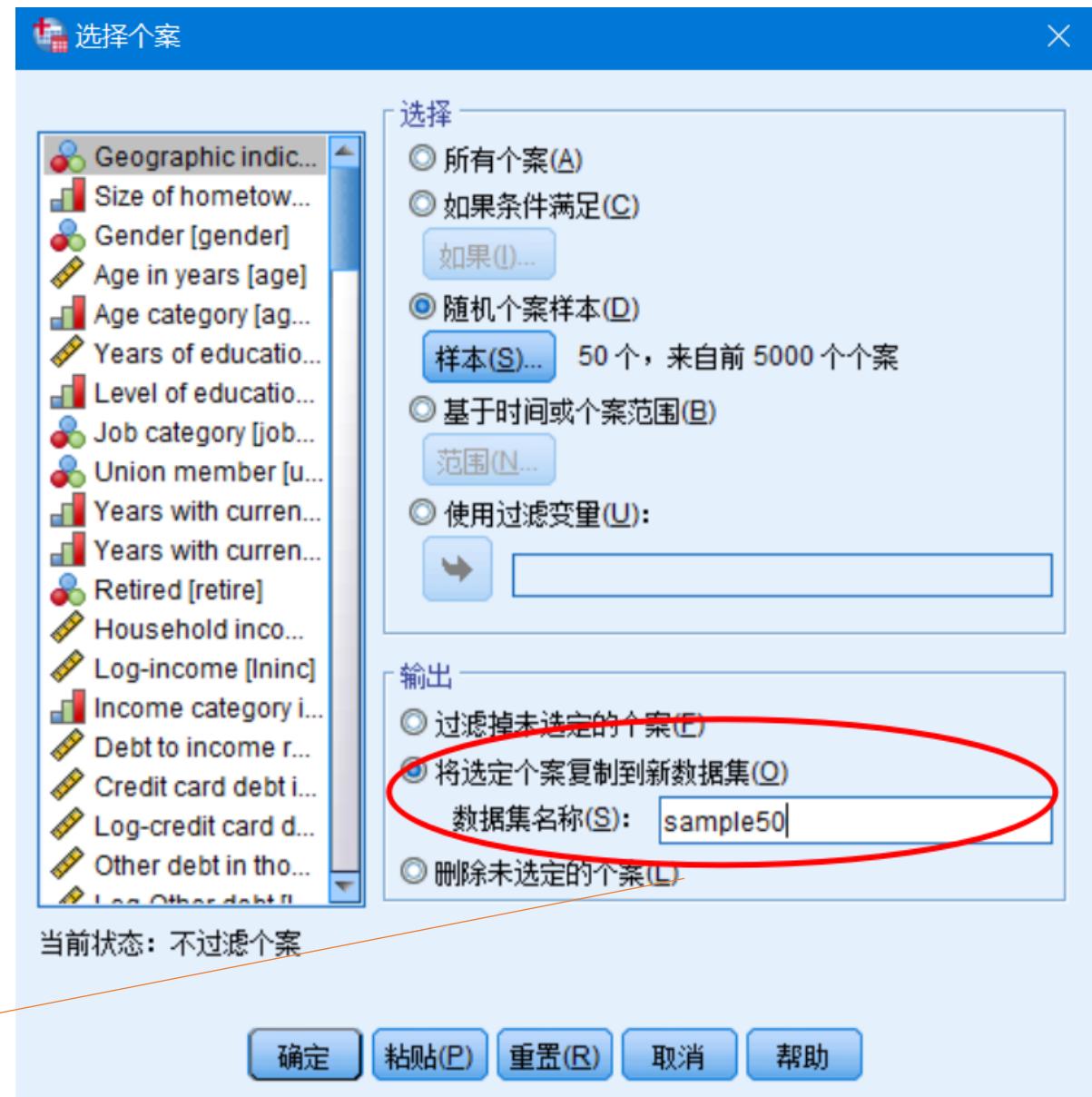
SPSS抽取随机样本

【样本】>输入需要选取的
样本数量及原
数据的样本量，
完成后点击
【继续】



SPSS抽取随机样本

在【输出】区域点击
【将选定个案复制到新
数据集(O)】，然后输
入新的数据集名称，完
成后点击【确定】

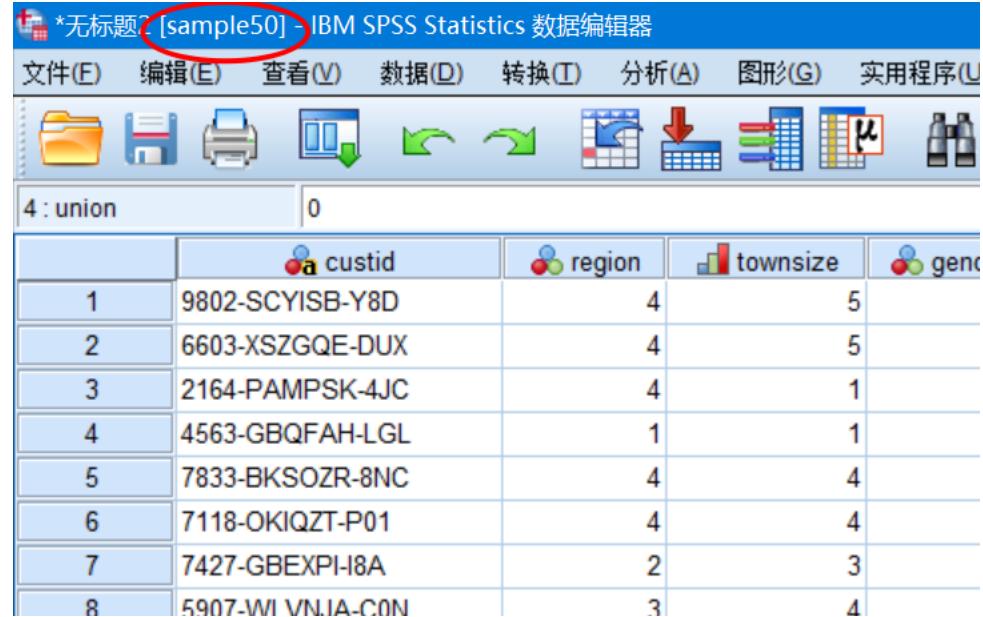


新的数据集的名称

SPSS抽取随机样本

以上操作完成后即可在新的窗口看到一个我们抽取出来的子样本。

请注意，这个数据集还没保存到我们的电脑上，如果需要保存数据，记得按Ctrl+S保存一下。



The screenshot shows the IBM SPSS Statistics Data Editor window. The title bar reads '*无标题 [sample50] - IBM SPSS Statistics 数据编辑器'. The menu bar includes: 文件(F), 编辑(E), 查看(V), 数据(D), 转换(T), 分析(A), 图形(G), 实用程序(U). The toolbar contains various icons for file operations and data analysis. The main data area displays a table titled '4 : union' with 8 rows and 5 columns. The columns are labeled: custid, region, townsize, and gender. The data is as follows:

	custid	region	townsize	gender
1	9802-SCYISB-Y8D	4	5	
2	6603-XSZGQE-DUX	4	5	
3	2164-PAMPSK-4JC	4	1	
4	4563-GBQFAH-LGL	1	1	
5	7833-BKSOZR-8NC	4	4	
6	7118-OKIQZT-P01	4	4	
7	7427-GBEXPI-I8A	2	3	
8	5907-WI VN.IA-C0N	3	4	

参考资料

- 《社会研究方法》（艾尔·巴比）第5、6章
- 《大众媒介研究导论》
- 《社会研究方法教程》（袁方）第十章
- 《社会研究方法》（风笑天）第5章
- 《心理与教育测量（第四版）》（戴海琦，张锋）

参考资料

- 《统计学的世界》 第2-4章
- 《社会研究方法》 第5章
- 开学了，如何在上课点名环节死里逃生？

<https://mp.weixin.qq.com/s/ambGtx-IJl4KXKuuTFXjdw>

谢谢！

