

带你走入统计学

统计与数据科学学院 马俊玲



“学者不能离开统计而究学，政治家不能离开统计而施政，事业家不能离开统计而执业。”

——经济学家和教育学家马寅初



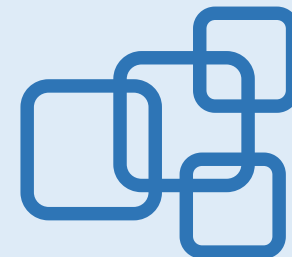
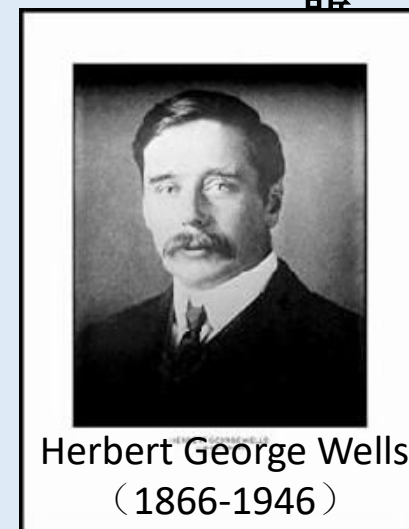


“和读写能力一样，终有一天统计思维会成为每个合格市民的必备素质。”
——科幻小说作家威尔斯

20世纪最著名的科幻作家威尔斯作品有《星球大战》、《时间机器》、《未来世界》、《隐身人》等，他还是一位社会改革家和预言家，撰写了大量关注现实，思考未来的作品。如《预测》、《世界史纲》等。

他曾预言装甲坦克、飞机在战争中的应用，他发明原子弹一词，描述原子弹爆炸，描述人类登月等。

他的《未来世界》（The Shape of Things to Come）一书1933年出版，书中有对日本全面侵华的预测。中译本1938年3月在中国出版。书中预测日本1935年就占领了北平和天津，但在1938年7月遭遇了瘟疫，后来国内又发生了粮荒、兵变和革命，此时美国参战，日本就崩溃了。



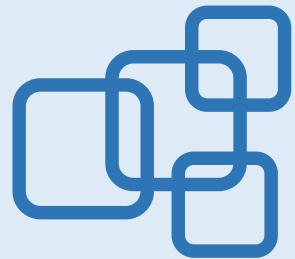
目录

Contents

- 01** 专题一——认识统计学
- 02** 专题二——做一次抽样调查
- 03** 专题三——从实践问题到统计量
- 04** 专题四——数据的初步分析
- 05** 专题五——数据分析的推断思维
- 06** 专题六——探索数据中的统计规律
- 07** 专题七——实证论文的形成



专题一 认识统计学



专题一 认识统计学

主要内容

- 一、什么是统计
- 二、数据
- 三、搜集数据的组织方式
- 四、统计软件

一、什么是统计？

(一) 统计学的定义

统计学：收集、分析、展示和解释数据的科学。

随机性中的统计规律

CCTV13新闻频道-每周质量报告 (来源: 央视网) 功能性服装质量调查



(二) 四个重要的术语

总体： 所研究的具有某些相同性质的全部单位或事件的整体。

样本： 亦可称为抽样总体, 是从总体中抽取部分单位所组成的整体, 用以分析总体。

参数： 亦可称为总体指标, 是综合测量的整个总体的某个数量特征。

统计量： 亦可称为样本指标, 是根据样本数据计算的综合测量值, 可用以反映或估计、推断总体的某个数量特征。

(三) 统计学的方法

1. 描述统计 关于搜集、展示一批数据，并反映这批数据特征的各种方法，其目的是为了正确地反映总体的数量特点。
2. 推断统计 根据样本统计量估计和推断总体参数的技术和方法。

二、变量和数据

变量——数据

(一) 数据的分类

1. 定性数据和定量数据

定性数据：用文字描述的。

如消费者对产品售后服务的总体评价等都属于文字描述的定性数据。

定量数据：用数量描述的。

如企业的净资产额、净利润额等。

2. 离散型数据和连续型数据

类型

离散型数据：数据只能取整数。
如一家公司的职工人数。

连续型变量的数据可以取介于两个数值之间的任意数值。
如销售额、经济增长率等。

3. 截面数据和时间序列数据

截面数据：所搜集的不同单位在同一时间的数据。例如，所有上市公司公布的2023年年度的净利润。

时间序列数据：所搜集的同一总体或单位在不同时间的数据。某公司公布的1993年到2023年的年度净利润就是时间序列数据。

面板数据

4. 原始数据和次级数据

原始数据：指直接从各个调查单位搜集的、尚未经过整理的统计数据资料，也称一手数据。

次级数据：指那些已经加工整理过的，往往是公开发表的数据，也称二手数据。

如从报纸杂志、统计年鉴、会计报表上取得的数据。

（二）数据的来源

1. 从政府机构、各种行业组织、公司和企业所公布的数据中获取。

就是把政府机构、各种组织和公司所公布的数据作为来源，这种数据往往是次级数据。

2. 设计一次试验以获取必要的数据。

例如，在检验洗衣机洗净程度的研究中，研究人员通过实际洗涤脏衣服，来研究哪种牌子的洗衣机效果最佳。

3. 从观察研究中获取。研究人员通常是在自然状态下，进行直接的观察。

例如，观察路口的交通流量、观察顾客在商场的购买行为和观察流水线上的产品质量等。

4. 进行一次调查。它对所调查人们的行为不进行任何控制，仅提出诸如出生年月、爱好、消费习惯、对某一事件的看法和其他特征方面的问题，然后对他们回答的结果进行整理、编码、列表和分析。

5. 其它，如网络平台数据。

三、数据搜集的组织方式

(一) 普查、抽样、统计报表制度和重点调查

1. 普查

特点: 工作量大, 时间性强, 需要大量人力和财力。

2. 抽样调查

特点: 1. 按随机原则从总体中抽取样本;

2. 以样本指标 (统计量)为依据推断总体参数或检验总体的某种假设;

3. 抽样误差可以事先计算并加以控制。

3. 统计报表制度

是按一定的表式和要求，自上而下统一布置，自下而上提供统计资料的一种统计调查方法。

4. 重点调查

特点：在总体中选择个别或部分重点单位进行调查。

重点单位指在总体中有举足轻重地位的单位，其标志值在总体标志总量中占有绝大比重。

5. 其它，如网络平台数据的搜集。

（二）抽样的优点

1. 适用的范围广。

对于有限总体，从理论上讲，既可以进行普查也可以进行抽样；

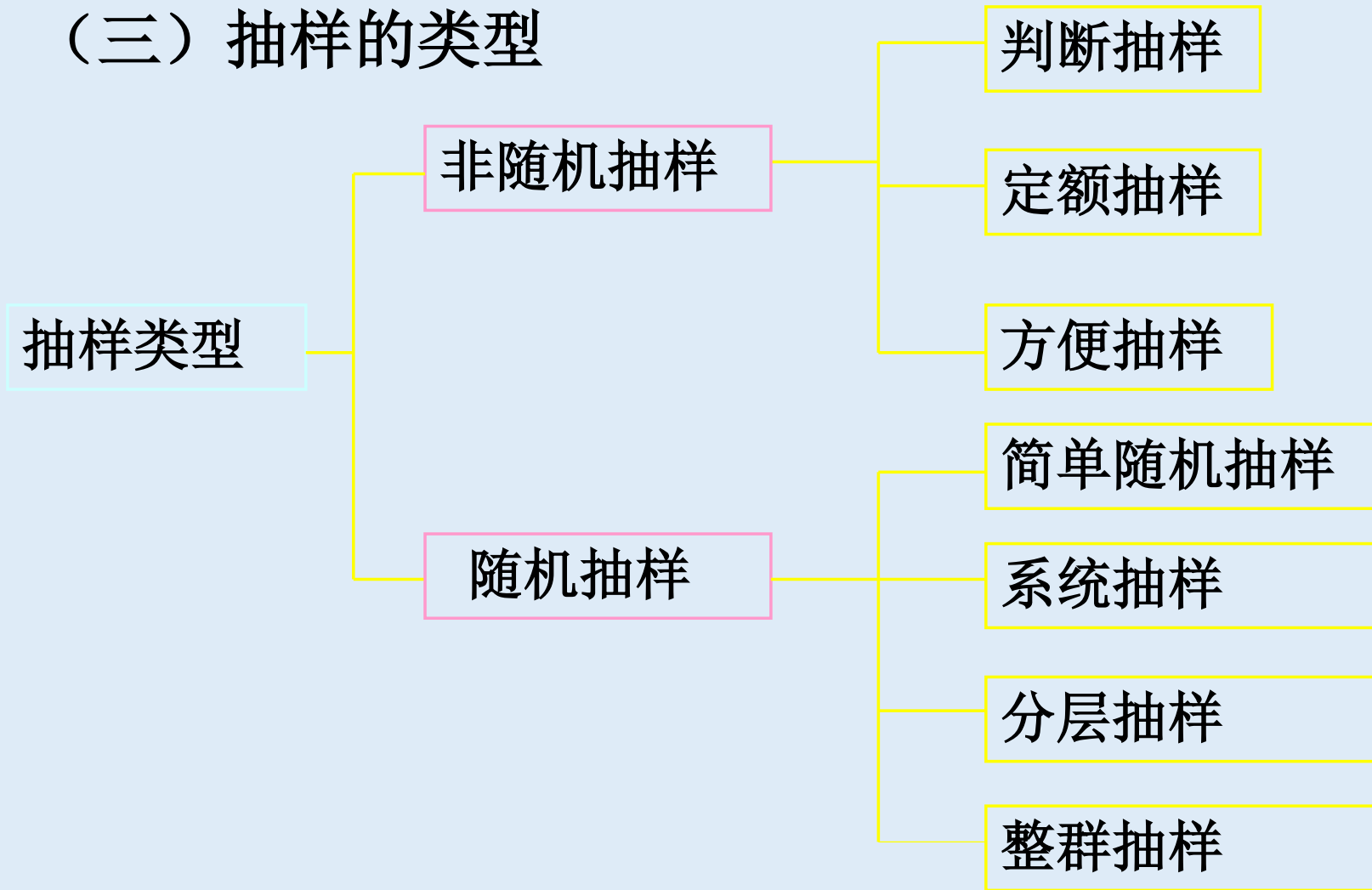
对于无限总体，就只能进行抽样。若理论上可以而实际上很难采用全面普查的情况，也只能采用抽样。

如产品质量的破坏性检验、居民住户调查等。

2. 与全面普查相比，抽样最大的优点是节省人、财、物力和时间。

3. 随机抽样可以比普查更为精确。

(三) 抽样的类型



非随机抽样

又称为非概率抽样,是不按照随机原则来抽取样本中的单位或个体。

特点

方便、快速和低成本

精确性差,结论缺乏普遍性

判断抽样

又称为典型调查,是从事有关工作的专家按照一定的标准有意识地在总体中选择若干有代表性的单位组成样本进行调查,代表单位的选取标准应根据统计研究的目的而定。

定额抽样

是根据已定的单位数抽取样本，往往是对总体了解甚少时采用。

如想获取某地区化妆品的销售情况，对该地区的5家商厦进行调查。

方便抽样

是为了取样方便，随意地抽取样本单位。

街头偶遇式调查就是一种最为常见的方便抽样。

随机抽样

——是根据随机原则来抽取样本单位。

简单随机抽样

——亦称为纯随机抽样

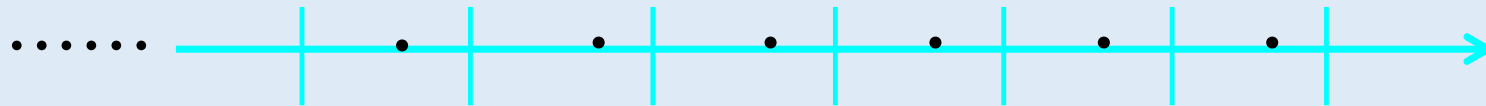
方法：在抽样框中的每个单位都具有相同的被抽中的机会，每个容量相同的样本被抽中的机会也是相同的。

抽取样本的方法：有放回抽样和无放回抽样。

系统抽样

——亦可称为等距抽样

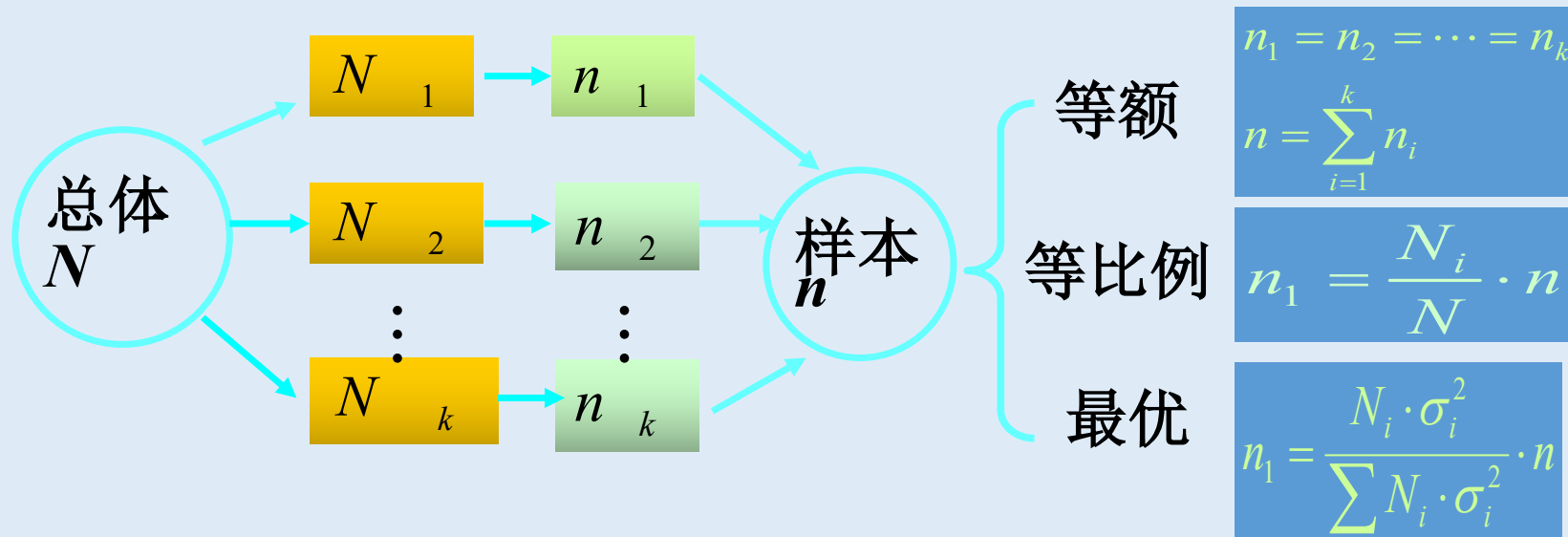
方法：抽样框中的 N 个单位被分成 k 个系统， k 等于抽样框的容量 N 除以所需的样本容量 n ，在抽样框中前面的 k 个个体或单位中随机抽出第一个样本单位，然后，可在其后的每隔 k 个单位抽取样本中其余的部分。



分层抽样

亦可称为类型抽样

方法：将总体全部单位分类，形成若干个类型组，后从各类型中分别抽取样本单位，合成样本。



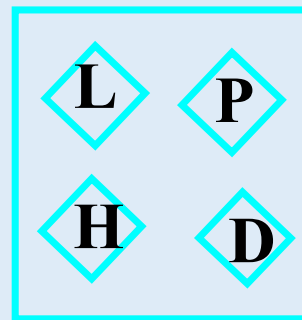
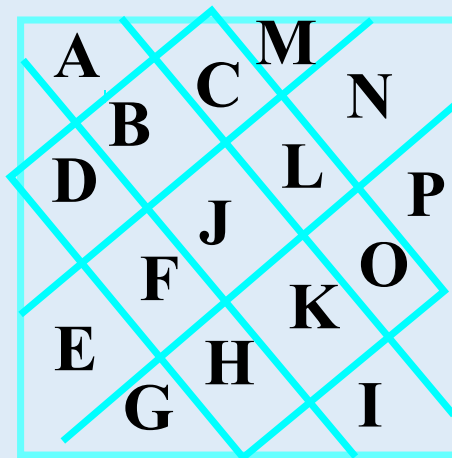
整群抽样

方法：首先把总体中的N个单位划分成为若干个群，并要求每个群对整个总体都具有代表性，然后对群进行简单随机抽样，并对抽中群内的所有单位进行调查研究。

总体群数 $R=16$

样本群数 $r=4$

样本容量



$$n = n_d + n_p + n_l + n_h$$

四、统计软件

(一) 统计软件的种类

- SPSS
- SAS
- R
- 其它：Excel（数据表格软件）、Eviews、Stata、Python、Matlab等

- SPSS:

这是一个很受欢迎的统计软件，它容易操作，输出漂亮，功能齐全，价格合理。对于非统计工作者是很好的选择。

- SAS:

这是功能非常齐全的软件，美国政府政策倾斜（“权威性”）许多美国公司使用。价格不菲，每年交费。即使赠送，条件苛刻。尽管现在已经尽量“傻瓜化”，仍然需要一定的训练才可以进入。

- R软件:

免费的, 志愿者管理的软件。编程方便, 语言灵活, 图形功能强大。有不断加入的各个方向统计学家编写的统计软件包。也可以自己加入自己算法的软件包。

这是发展最快的软件, 受到世界上统计师生的欢迎。是用户量增加最快的统计软件。

对于一般非统计工作者来说, 主要问题是它没有“傻瓜化”。

（二）统计软件使用中的问题

- 统计软件傻瓜式功能的发展，使得数据分析变得如此简单，只要输入数据，点几下鼠标，做一些选项，马上就得到令人惊叹的漂亮结果。
- 是否傻瓜式的统计软件使用可以代替统计课程了？

（三）SPSS软件初识

顾客售后满意度调研问卷（节选）

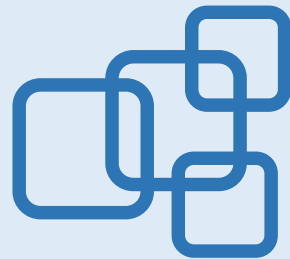
- S2 请问您的年龄是（单选）
 1. 18岁以下（终止访问）
 2. 18~39岁
 3. 40~60岁
 4. 60岁以上（终止访问）
- S3 记录顾客的性别（单选）
 1. 男性
 2. 女性
- Q1 请问您有多大可能向您的家人或朋友推荐×××商店？（1~10分）
[10分代表肯定会推荐，0分代表肯定不会推荐]
- Q9 在过去的3个月中，平均来讲，您到×××商店购物的频次是？（单选）
 1. 每天都来
 2. 每周5~6次
 3. 每周3~4次
 4. 每周1~2次
 5. 每月2~3次
 6. 每月1次或以下
- Q10 请问您通常采用何种交通方式去×××商店购物？（最多限三项）
 1. 步行
 2. 公交车
 3. 自行车
 4. 自有摩托车
 5. 自己开车
 6. 出租车（打的）
 7. 商场的免费购物班车
 8. 出租摩托车
 9. 人力三轮车
 10. 电瓶车/电动车
 11. 地铁
 12. 轻轨
 13. 其他方式（请具体说明）

来自数据集：SPSS数据例
顾客满意度.sav



思考与练习

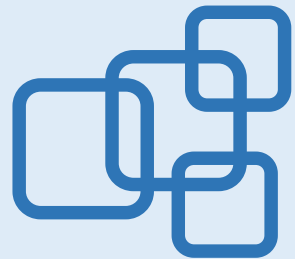
1. 准备好SPSS软件，练习录入数据、读取数据等基本操作。
2. 就我校各个方面的情况进行思考，发现一两个感兴趣的具有研究意义的问题。
3. 登录我校图书馆数据库网页看看都有哪些变量的数据。
4. 登录国家统计局网页找找国家都发布哪些数据。





参考资料

- [1] 王静龙.梁小筠.魅力统计.中国统计出版社,2012
- [2] 张文彤, 钟云飞编著, SPSS数据分析与挖掘实战案例精粹, 清华大学出版社, 2013.
- [3] 吴喜之编著, 统计学: 从数据到结论, 第四版, 中国统计出版社, 2019.





THANK YOU!

