

1



2

第1章 统计学概念

- 1.1 决策与统计思考
- 1.2 什么是统计学?
- 1.3 总体与样本
- 1.4 抽样的概念
- 1.5 数据的种类
- 1.6 描述统计学与推断统计学

3/18

3

1.1 决策与统计思考



- 为解决某一问题而进行的决策过程
 - ① 整理核心问题
 - ② 选定调查对象
 - ③ 明确所需数据、制定实验/调查方法
 - ④ 收集、整理数据、统计分析 → 信息
 - ⑤ 发现重要因素(factor)
 - ⑥ 开发统计模型、估计
 - ⑦ 拟合优度检验、模型最优化 → 解决方案
 - ⑧ 得出结论、决策建议

4/18

4

1.2 什么是统计学?



[定义 1-1] 统计学(statistics)

- 在不确定的情况下，以数据为依据，有助作出科学性决策的理论和
方法体系。

- 语源
 - 拉丁语 statisticus(概率), statisticum(状态)
 - 意大利语 statista(国家、政治家)

5/18

5

1.2.2 统计学的应用领域



- 数据的收集、分类、表达、分析、推断、解读
- 工学、农业、生命科学、经济学、经营学、产业研究、质量
保证、市场调查、舆论调查 ...
- 统计学者的角色
 - 所需数据的形态
 - 收集数据的方法
 - 为寻找问题最佳答案而使用的分析方法

6/18

6

1.2.3 统计学的研究领域

- 开发统计理论与方法论
 - 概率论、概率过程论、数理统计、渐进式推断
 - 线性/非线性模型、回归分析、实验策划、统计计算
 - 贝叶斯推断、非参数推断、时间序列分析等
- 统计方法论应用
 - 生物统计、生命信息学、品质管理、可信性分析
 - 工学统计、计量经济学、金融统计、保险统计、气象统计

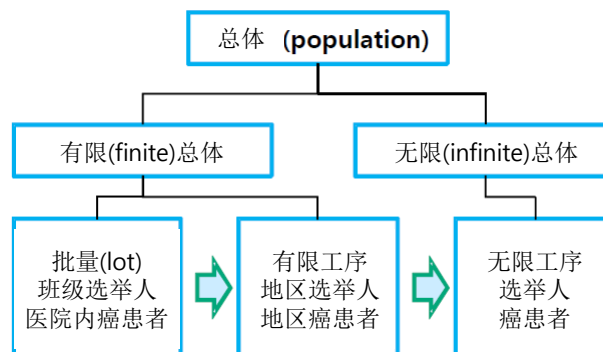
7/18

7

1.3 总体与样本

[定义 1-2] 总体(population)

- 作为统计分析研究对象的所有个体的集合，为作出正确的决策而需加以关注与研究的对象



8/18

8

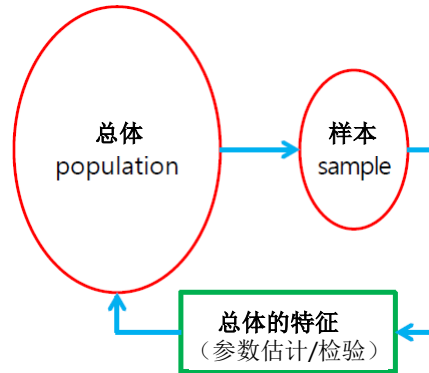
1.3 总体与样本

[定义 1-3] 参数(parameter)

- 作为体现总体的特征的数值，只有是为了作出正确的决策而需要特别关注的参数才可以成为研究的对象。
- ✓ 均值、方差、标准差、比例等

[定义 1-4] 样本(sample)

- 为了把握总体的特征，根据一定的规则而从总体中抽取的部分集合。



9/18

9

1.3 总体与样本

• 统计推断(inference)

- 推测总体的特征(均值、方差、比例等)
估计(estimation)、假设检验(hypothesis test)

• 统计量(statistic)

- 为了推测总体的特征(均值、方差、比例等)而使用的样本函数

• 估计量(estimator)

- 估计参数时所使用的统计量
样本均值、样本方差、样本比例等

$$\begin{aligned}\bar{X} &\Rightarrow \mu \\ S^2 &\Rightarrow \sigma^2 \\ \frac{X}{n} &\Rightarrow p\end{aligned}$$

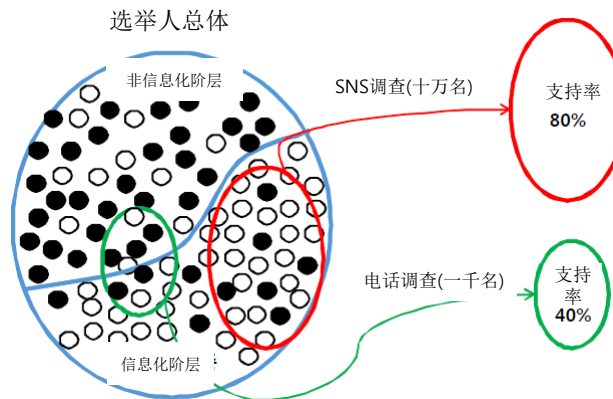
10/18

10

1.4 抽样的概念

[定义 1-5] 抽样(sampling)

- 为了对总体作出正确的决策而从总体中抽取样本的一系列过程



11/18

11

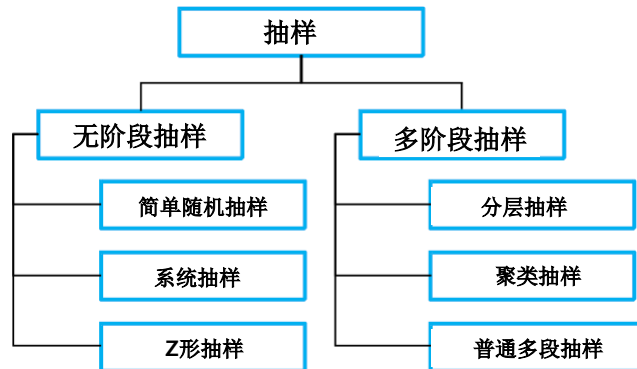
1.4.1 抽样的必要性

- 样本调查(sample survey)
 - 在总体中抽取样本的抽样调查方式
- 抽样的必要性
 - (1) 现实性制约
 - ✓ 无限总体、无意义、破坏性检查 ...
 - (2) 费用节省
 - ✓ 人工费、分析费用、附加费用 ...
 - (3) 时间节省
 - ✓ 迅速的调查、时间上的制约 ...
 - (4) 品质层面
 - ✓ 疲劳度考虑、检查失误 ...

12/18

12

1.4.2 抽样方式



[定义 1-6] 随机样本(random sample)

- 从总体中随机抽样后得出的独立并具有相同分布的样本

13/18

13

1.4.2 抽样方式

- 从20个箱子(每个箱子中50个)中抽取100样本

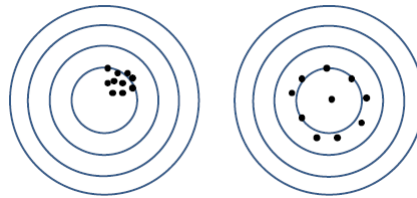
- ① 简单随机抽样 1~1000之间随机选择100个
- ② 系统抽样 1~10 之间随机选择一个 → 7
→ 7, 17, 27, ..., 977, 987, 997等选取100个
- ③ Z形抽样 → 7, 14, 27, ..., 974, 987, 994等选取100个
- ④ 2段抽样 随机抽取5个箱子, 然后从每个箱子中随机抽取20个
- ⑤ 分层抽样 从20个箱子中各随机抽取5个
- ⑥ 聚类抽样 随机抽取2个箱子, 每个箱内的50个全部选取

14/18

14

1.4.3 抽样与误差

- 误差(error)：参数的实际值与估计值之间的差异
- 误差的类型 $x = \mu + s + r + m$
 - 估计误差：估计器械的不准确、估计技术不足
 - 分析误差：分析装备的不准确、分析技术不足
 - 抽样误差：倾斜抽样
 - 缩放误差：在大样本中抽取用于估计的样本时出现倾斜
- 误差的探讨
 - ① 可信性(reliability)
 - ② 精密度(precision)
 - ③ 准确性(accuracy)



15/18

15

1.5 数据的种类

- 数据尺度(Scale)类型

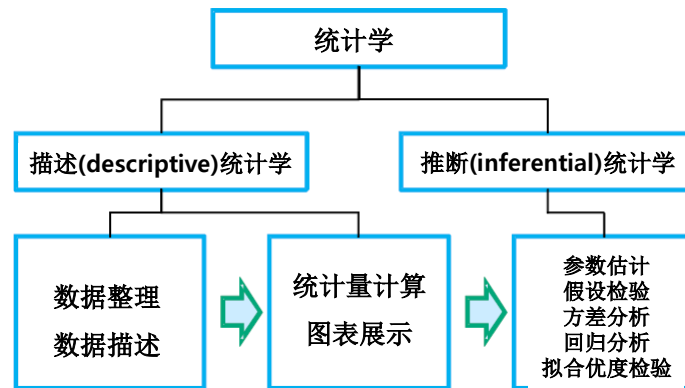
尺度类型	特征	实例
名目尺度 (nominal)	没有顺序、大小等意义	教科编号、车牌号、性别、身份证号末尾数、彩票号码
顺序尺度 (ordinal)	具有顺序意义，但没有间隔意义	职位、阶级、优先顺序、偏好度
区间尺度 (interval)	具有顺序和间隔意义，但没有比例(或者绝对0)意义	待机号码、华氏温度、智能指数、标准分数
比例尺度 (ratio)	具有顺序、间隔、比例(或者绝对0)等意义	浓度、重量、长度、压力、速度、支持率

16/18

16

1.6 描述统计学与推断统计学

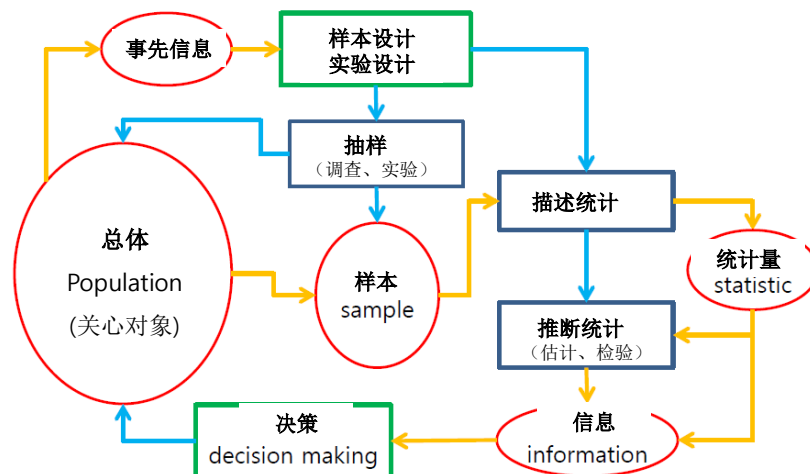
- 统计学的构成体系



17/18

17

- 统计决策过程



18/18

18