

9.1 随机抽样



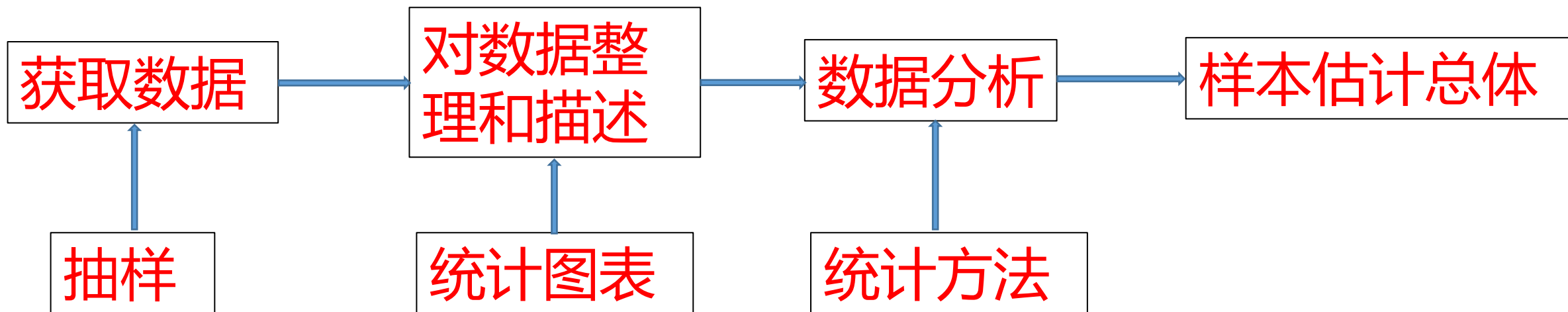
学习目标

1. 正确理解简单随机抽样的概念.
2. 掌握抽签法及随机数法的步骤.

引入

人口总量、经济增长率、就业状况、物价指数、产品合格率、商品的销售额、农作物的产量、人均水资源、居民人均年收入、电视台节目收视率、学生平均身高等统计数据。 **统计学**

统计学是通过收集数据和分析数据来认识未知现象的一门科学。



2010年我国进行了第六次人口普查，对全国人口普遍地、逐户逐人地进行一次性调查登记.调查内容包括每位居民的姓名、性别、年龄、民族、受教育程度等.居民是调查对象，而居民的性别、年龄、民族、受教育程度等是要调查的指标.由于不同调查对象的指标值往往不同，它是一个变化的量，所以常把指标称为变量.

对每一个调查对象都进行调查的方法，称为全面普查**，又称**普查**.**

调查对象的全体称为总体**.**

组成总体的每一个调查对象称为**个体**.

为了强调调查的目的，也可以把调查对象的某些指标的全体作为总体，每一个调查对象的相应指标作为个体.

根据一定目的，从总体中抽取一部分个体进行调查，并以此为依据对总体的情况作出估计和推断的调查方法，称为**抽样调查**.

从总体中抽取的那部分个体称为**样本**.

样本中包含的个体数称为**样本量**.

调查样本获得的变量值称为**样本的观测数据**，简称**样本数据**.

问题1 : 为了了解全国高中生的视力情况, 需要将全中国所有高中生逐一进行检查吗?

容量大!

问题2 : 要检查某超市销售的牛奶含菌量是否合格, 需要将该超市的所有牛奶的包装袋都打开逐一检查吗?

有破坏性!

要了解全国高中生的视力情况：你认为哪种调查方式较适合？

(1) 对全国所有的高中生进行视力测试；

属于普查，工作量太大，不方便，没有必要

(2) 对某一所著名中学的高中生进行视力测试；

这种方法缺乏普遍性，不合适。

(3) 在全国按东、南、西、北、中分片，每个区域各抽 3 所中学，对这 15 所中学的全部高中生进行视力测试。

这种调查具有可操作性及代表性。

人们在研究某个自然现象或社会现象时，会遇到不方便、不可能或不必要对所有对象作调查的情况，往往采用抽样调查的方法。

为什么要用样本的情况估计总体的相应情况？

我们要考察的总体的个体数往往很多，且有时虽然总体中的个体数不是很多，但考查时带有破坏性。所以通常总体中抽取一个样本，通过样本来研究总体。则抽取的样本数据就必须能够很好地反映总体的情况，即样本含有和总体基本相同的信息。

如何抽取样本？怎样使抽取的样本充分地反映总体的情况？

合理、公平、有代表性

两种抽样方法---简单随机抽样和分层随机抽样。

请同学们阅读课本P174探究以下内容.并回答下列问题:

(1) 简单随机抽样有几种类型? 如何定义? 各有什么特点?

一般地, 设一个总体有 N (N 为正整数) 个个体, 从中逐个抽取 n ($1 \leq n < N$) 个个体作为样本, 如果抽取是放回的, 且每次抽取时总体内的各个个体被抽到的概率都相等, 我们把这样的抽样方法叫做**放回简单随机抽样**; 如果抽取是不放回的, 且每次抽取时总体内未进入样本的的各个个体被抽到的概率都相等, 我们把这样的抽样方法叫做**不放回简单随机抽样**.

放回简单随机抽样和**不放回简单随机抽样**统称为**简单随机抽样**.

通过简单随机抽样获得的样本称为简单随机样本.

有放回的摸球中，同一个小球有可能被摸中多次，极端情况是每次摸到同一个小球，而被重复摸中的小球只能提供同一个小球的颜色信息.不放回的摸球中，每次摸球都在余下的球中随机摸取，就可以避免同一球被重复摸中.当把球都摸出时，就能完全了解了红球的比例，而有放回的摸球一般不能对袋中红球的比例做出准确的判断.

说明:如没特殊说明，本章所称简单随机抽样指不放回简单随机抽样.

请同学们阅读课本P175 ~ 176，问题1以后内容.并回答下列问题：

简单随机抽样的方法有几种？各有什么特点？

1. 简单随机抽样方法——抽签法（抓阄法）

步骤：

(1) 编号制签（从0到N-1或1到N编号）

(2) 搅拌均匀

(3) 逐个不放回抽取n次

例如：我们班有44名学生，现从中抽出5名学生去参加学生座谈会，要使每名学生的机会均等，我们应该怎么做？

1. 简单随机抽样方法——抽签法（抓阄法）

• 优点：

简单易行，当总体的个数不多时，抽签法能够保证每个个体被抽中的机会都相等

• 缺点：

1) 当总体的个数非常大时，制作号签费时费力；

2) 号签很多时较难搅拌均匀，难以保证每个个体入选的可能性相等，从而造成代表性差。

2. 简单随机抽样方法——随机数法

步骤：（1）将总体中的 N 个个体编号；

（2）用随机数工具产生 $1 \sim N$ 范围内的整数随机数，把产生的随机数作为抽中的编号；

（3）使与编号对应的学生进入样本；

（4）重复上述过程，直到抽足样本所需要的人数。

说明：如果生成的随机数有重复，即同一编号多次被抽到，可以剔除重复的编号并重新产生随机数，直到产生不同的编号个数等于样本所需要的人数。

方法：（1）用随机试验生成随机数

①准备10个大小、质地一样的小球，小球上分别写上数字0,1,2, ..., 9, 把它们放入一个不透明的袋子.②总体中个体编号为n位数，就从袋中有放回地摸取n次，每次摸取前充分搅拌，并把第1,2, ..., n次摸到的数字分别作为编号的相应位数上的数字，这样就产生了一个随机数.③如果这个随机数在1~N的范围内，就代表对应编号的个体被抽中，否则舍弃编号；④重复这个过程，直到得到n个编号.

（2）用信息技术生成随机数

①用计算器生成随机数

②用电子表格软件生成随机数

③用R统计软件生成随机数

思考：在简单随机抽样中，样本量是否越大越好？ **不一定**
[微判断]

1. 抽签法和随机数法都适用于总体容量和样本容量较小时的抽样。()

2. 用随机数工具生成的随机数有重复，需剔除重复的编号。()

3. 样本量的大小不影响样本的代表性。()

提示 容量大的样本的代表性比容量小的样本的代表性好，尤其是样本量不大时，增加样本量可以较好地提高估计的效果。

[微训练]1.下面抽样方法是简单随机抽样的是()

A.从平面直角坐标系中抽取5个点作为样本

B.可口可乐公司从仓库中的1 000箱可乐中一次性抽取20箱进行质量检查

C.某连队从200名战士中，挑选出50名最优秀的战士去参加抢险救灾工作

D.从10个手机中逐个不放回地随机抽取2个进行质量检验(假设10个手机已编好号，对编号随机抽取)

解析 A中，平面直角坐标系中有无数个点，这与要求总体中的个体数有限不相符，故错误；C中，50名战士是最优秀的，不符合简单随机抽样的等可能性，故错误.

答案 B、D

题型一 简单随机抽样的概念理解

【例1】 (1)关于简单随机抽样的特点有以下几种说法，其中不正确的是()

A.要求总体中的个体数有限

B.从总体中逐个抽取

C.这是一种不放回抽样

D.每个个体被抽到的机会不一样，与先后顺序有关

(2)下列问题中最适合用简单随机抽样方法的是()

A.某学校有学生1 320人，卫生部门为了了解学生身体发育情况，准备从中抽取一个容量为300的样本

B.为了准备省政协会议，某政协委员计划从1 135个村庄中抽取50个进行收入调查

C.从全班30名学生中，任意选取5名进行家访

D.为了解某地区癌症的发病情况，从该地区的5 000人中抽取200人进行统计

解析 (1)简单随机抽样，除具有A，B，C三个特点外，还具有等可能性，每个个体被抽取的机会相等，与先后顺序无关.

(2)A中不同年级的学生身体发育情况差别较大，B，D的总体容量较大，C的总体容量较小，适宜用简单随机抽样.

答案 (1)D (2)C

规律方法 1.一个抽样试验能否用抽签法，关键看两点：一是制签是否方便；二是个体之间差异不明显.

2.应用抽签法时应注意以下几点：

(1)编号时，如果已有编号可不必重新编号；

(2)号签要求大小、形状完全相同；

(3)号签要均匀搅拌；

(4)根据实际需要采用有放回或无放回抽取.

素养落地

- 1.通过学习简单随机抽样的相关概念以及抽签法和随机数法的抽样步骤，重点培养数据分析的核心素养.
- 2.要判断所给的抽样方法是不是简单随机抽样，关键是看它们是否符合简单随机抽样的定义，即简单随机抽样的四个特点：总体有限、逐个抽取、无放回抽取、等可能抽取.
- 3.当总体容量和样本容量都不大时，用抽签法抽样；当总体容量较大，样本容量不大时，用随机数法抽样.

总体中有 N 个个体，它们的变量值分别为 Y_1, Y_2, \dots, Y_N ，则称

$$\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{1}{N} \sum_{i=1}^N Y_i$$

为总体均值，又称总体平均数。

如果总体的 N 个变量值中，不同的值有 k ($k \leq N$)个，不妨记为 Y_1, Y_2, \dots, Y_k ，其中 Y_i 出现的频数 f_i ($i = 1, 2, \dots, k$)，则总体均值还可以写成加权平均数的形式

$$\bar{Y} = \frac{Y_1 f_1 + Y_2 f_2 + \dots + Y_k f_k}{f_1 + f_2 + \dots + f_k} = \frac{1}{N} \sum_{i=1}^k f_i Y_i$$

$$(N = f_1 + f_2 + \dots + f_k)$$

总体的一个容量为 n 的样本，它们的变量值分别为 y_1, y_2, \dots, y_n ，则称

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{n} \sum_{i=1}^n y_i$$

为样本均值，又称样本平均数.在简单随机抽样中，我们常用样本平均数 \bar{y} 去估计总体平均数 \bar{Y} .

练习：用抽签法抽取的一个容量为5的样本，它们的变量值分别为2, 3, 5, 7, 9，则该样本的平均数为(C)

A.4.5

B.4.8

C.5.2

D.6

某学校抽取100位老师的年龄，得到如下数据：

年龄(单位：岁)	32	34	38	40	42	43	45	46	48
频数	2	4	20	20	26	10	8	6	4

估计这个学校老师的平均年龄.

解 $\bar{y} = \frac{1}{100} \times (32 \times 2 + 34 \times 4 + 38 \times 20 + 40 \times 20 + 42 \times 26 + 43 \times 10 + 45 \times 8 + 46 \times 6 + 48 \times 4) = 41.1$ (岁)，即这个学校老师的平均年龄约为 41.1 岁.