

1. 用系统抽样法从 160 个灯泡中抽取一个容量为 20 的样本。若将这 160 个灯泡编号为  $1 \sim 160$ ，若在第 16 个被抽中的个体编号为 126，起始抽取编号为多少？

**解** 抽样间距： $k = \frac{N}{n} = \frac{160}{20} = 8$ ，

设起始抽样编号为  $r$ ，则：其余抽样编号为： $r, r+k, r+2k, \dots, r+15k, \dots$ ，

因此第 16 个为： $r + 15k = 126$

所以起始抽样编号为： $r = \frac{126}{15 \div 8} = 6$

4. 设总体有 14 个个体，并按照  $1 \sim 14$  进行编号。欲要以系统抽样法抽取容量为  $n = 4$  的样本，且第一个抽中的编号为 1，则其余 3 个样本编号依次为多少？

**解** 由题意可以，采取系统抽样抽取样本；又因抽样间距  $k$  不为整数，因此采用圆形等距抽样。

$k = \frac{N}{n} = 3.5$ ，取与之最近的整数  $k=4$ ，因为第一个编号为 1，则其余三个编号为  $1+4=5$ ， $5+4=9$ ， $9+4=13$ 。

5. 已知 A，B，C 三个车间一天内生产的产品分别是 280 件、95 件、125 件，为了掌握各车间产品质量情况，需从中抽取一个样本容量为 100 的样本，应该如何抽取？

**解** 由题意可知，产品已分为三层，由 A，B，C 车间生产，因此采用分层抽样。

样本容量和总体容量之比为  $100 : 280 + 95 + 125 = 1 : 5$

利用抽样比确定每个车间应该抽取的产品数，依次为： $280/5 = 56$   $95/5 = 19$   $125/5 = 25$ ，

最后，利用简单随机抽样或系统抽样方法，从三个车间中分别抽取 56，19，25 个产品，合在一起后即为我们所抽取的样本。

8. 某校 500 名学生中, 有 200 人的血型为 O 型, 有 125 人的血型为 A 型, 有 125 人的血型为 B 型, 有 50 人的血型为 AB 型。为了研究血型与色弱的关系, 需从中抽取一个容量为 20 的样本, 怎样抽取样本比较合适? 请写出具体抽样过程。

**解** 因为总体由差异明显的不同血型组成, 因此采用分层抽样;

法一: 样本容量和总体容量之比为  $20/500 = 1:25$

利用抽样比确定每个血型应该抽取的人数, 依次为:  $200/25 = 8$   $125/25 = 5$   $125/25 = 5$   $50/25 = 2$ ;

最后, 利用简单随机抽样或系统抽样方法, 从四个血型中分别抽取 8, 5, 5, 2 个学生, 合在一起后即为我们所抽取的样本。

法二: 因为需要研究血型和色弱的关系, 可以采用分层抽样中的等额样本, 则:  $\frac{20}{4} = 5$ ;

因此利用简单随机抽样或系统抽样方法, 从四个血型中分别抽取 5 个学生, 合在一起后即为我们所抽取的样本。

13. Google 曾经有一道非常经典的面试题: 给定一个长度为  $N$  的链表, 其中  $N$  很大, 而且不知道确切知道  $N$  的大小。你的任务是从这  $N$  个元素中随机取出  $k$  个元素, 当只能遍历这个链表一次, 而且必须保证取出的元素恰好是  $k$  个, 且它们是完全随机的 (每个元素被抽中的概率相等)?

**解** 采样水库抽样就可实现题目中的要求:

(1) 首先将长度为  $N$  的链表中的前  $k$  个元素保留下来, 构建一个大小为  $k$  的水库;

(2) 对于链表中的第  $m$  个元素 ( $m > k$ ), 以  $\frac{k}{m}$  的概率决定是否由这个元素替换掉水库中的一个元素;

(3) 循环步骤 (2) 直到链表末尾。

14. 假设有一个大小为  $N$  的数组，其中未知  $N$  的值，依次扫描该数组的每个元素，为每个元素赋予一个随机数，然后使用 Top-k（譬如最大  $k$  个数）得到需要的  $k$  个元素。请问该方法是一种等概率抽样方法吗？

**解** 该方法是一个等概率抽样方法；对于前  $k$  个元素，会被选入 Top-k，概率为  $\frac{k}{k} = 1$ ；

设随机数范围为  $[a, b]$ ，则 Top- $k$  的范围为  $[a + \frac{k}{m}(b - a), b]$ ，则随机数在 Top- $k$  区间的概率为：

$$\frac{b - (a + \frac{k}{m}(b - a))}{b - a} = \frac{k}{m}$$

对于第  $m$  个元素 ( $m > k$ )，对它赋予随机数，那么它有  $\frac{k}{m}$  的概率被选入前  $k$  个数，每个元素被抽取的概率相等。