

2.1.2 系统抽样

探究

某学校为了了解高一年级学生对教师教学的意见,打算从高一年级 500 名学生中抽取 50 名进行调查.除了用简单随机抽样获取样本外,你能否设计其他抽取样本的方法?

我们按照这样的方法来抽样:首先将这 500 名学生从 1 开始进行编号,然后按号码顺序以一定的间隔进行抽取.由于 $\frac{500}{50}=10$,所以抽取的两个相邻号码之差可定为 10,即从 1~10 中随机抽取一个号码,例如抽到的是 6 号,每次增加 10,得到

6, 16, 26, 36, ..., 496.

这样我们就得到一个容量为 50 的样本.这种抽样方法是一种**系统抽样**(systematic sampling).

一般地,假设要从容量为 N 的总体中抽取容量为 n 的样本,我们可以按下列步骤进行系统抽样:

(1) 先将总体的 N 个个体编号.有时可直接利用个体自身所带的号码,如学号、准考证号、门牌号等;

(2) 确定分段间隔 k ,对编号进行分段.当 $\frac{N}{n}$ (n 是样本容量)是整数时,取 $k=\frac{N}{n}$;

(3) 在第 1 段用简单随机抽样确定第一个个体编号 l ($l \leq k$);

(4) 按照一定的规则抽取样本.通常是将 l 加上间隔 k 得到第 2 个个体编号 $(l+k)$,再加 k 得到第 3 个个体编号 $(l+2k)$,依次进行下去,直到获取整个样本.

请将这种抽样方法与简单随机抽样做一个比较,你认为这种抽样方法能提高样本的代表性吗?为什么?

如果遇到 $\frac{N}{n}$ 不是整数的情况,可以先从总体中随机地剔除几个个体,使得总体中剩余的个体数能被样本容量整除.

练习

1. 你认为系统抽样有哪些优点和缺点?
2. 设某校共有 118 名教师,为了支援西部的教育事业,现要从中随机地抽出 16 名教师组成暑期西部讲师团.请用系统抽样法选出讲师团成员.
3. 有人说,我可以借用居民身份证号码(18 位)来进行中央电视台春节联欢晚会的收视率调查:在 1~999 中抽取一个随机数,比如这个数是 632,那么身份证后三位数是 632 的观众就是我要调查的对象.请问,这样所获得的样本有代表性吗?为什么?



广告中数据的可靠性

今天已进入数字时代,各种各样的统计数字和图表充斥着媒体,由于数字给人的印象直观具体,所以让数据说话是许多广告的常用手法.但广告中的数据可靠吗?

在各类广告中,你会经常遇到由“方便样本(即样本没有代表性)”所产生的结论.例如,某减肥药的广告称,其减肥的有效率为 75%.见到这样的广告你会怎么想?通过学习统计这部分内容,你会提出下面的问题吗?这个数据是如何得到的;该药在多少人身上做过试验,即样本容量是多少;样本是如何选取的;等等.假设该药仅在 4 个人身上做过试验,样本容量为 4,用这样小的样本量来推断总体是不可信的.

“现代研究证明,99%以上的人感染有螨虫……”这是一家化妆品公司的广告.第一次听到此话的人会下意识地摸一下自己的皮肤,甚至会感觉到有虫在里面蠕动,恨不得立即弄些药膏抹抹,广告的威慑作用不言而喻.但这里 99%是怎么得到的?研究共检测了多少人?这些人是如何挑选的?如果检测的人都是去医院看皮肤病的人,这个数据就不适用于一般人群.

某化妆品的广告声称:“它含有某种成分可以彻底地清除脸部皱纹,只需 10 天,就能让肌肤得到改善.”我们看到的数字很精确,而“能让肌肤得到改善”却是很模糊的.这样的数字能相信吗?试验是在什么样的皮肤上做的?试验的人数是多少?

当我们见到广告中的数据时一定要多提几个问题.