

微专题 95 高中涉及的统计学知识

一、基础知识：

（一）随机抽样：

1、抽签法：把总体中的 N 个个体编号，把号码写在号签上，将号签放在一个容器中搅拌均匀后，每次从中抽取一个号签，连续抽取 n 次，就得到容量为 n 的样本

2、系统抽样：也称为等间隔抽样，大致分为以下几个步骤：

（1）先将总体的 N 个个体编号

（2）确定分段间隔 k ，设样本容量为 n ，若 $\frac{N}{n}$ 为整数，则 $k = \frac{N}{n}$

（3）在第一段中用简单随机抽样确定第一个个体编号 l ，则后面每段所确定的个体编号与前一段确定的个体编号差距为 k ，例如：第 2 段所确定的个体编号为 $l+k$ ，第 m 段所确定的个体编号为 $l+(m-1)k$ ，直至完成样本

注：（1）若 $\frac{N}{n}$ 不是整数，则先用简单随机抽样剔除若干个个体，使得剩下的个体数能被 n 整除，再进行系统抽样。例如 501 名学生所抽取的样本容量为 10，则先随机抽去 1 个，剩下的 500 个个体参加系统抽样

（2）利用系统抽样所抽出的个体编号排成等差数列，其公差为 k

3、分层抽样：也称为按比例抽样，是指在抽样时，将总体分成互不交叉的层，然后按照一定的比例，从各层独立地抽取一定数量的个体，将各层取出的个体合在一起作为样本。

分层抽样后样本中各层的比例与总体中各个层次的比例相等，这条结论会经常用到

（二）频率分布直方图：

1、频数与频率

（1）频数：指一组数据中个别数据重复出现的次数或一组数据在某个确定的范围内出现的数据的个数。

（2）频率：是频数与数据组中所含数据的个数的比，即频率=频数/总数

（3）各试验结果的频率之和等于 1

2、频率分布直方图：若要统计每个小组数据在样本容量所占比例大小，则可通过频率分布表（表格形式）和频率分布直方图（图像形式）直观的列出

（1）极差：一组数据中最大值与最小值的差

（2）组距：将一组数据平均分成若干组（通常 5-12 组），则组内数据的极差称为组距，所以

有组距=极差/组数

(3) 统计每组的频数，计算出每组的频率，便可根据频率作出频率分布直方图

(4) 在频率分布直方图中：横轴按组距分段，纵轴为“频率/组距”

(5) 频率分布直方图的特点：

① 频率 = $\frac{\text{频率}}{\text{组距}} \times \text{组距}$ ，即分布图中每个小矩形的面积

② 因为各试验结果的频率之和等于 1，所以可得在频率分布直方图中，各个矩形的面积和为 1

(三) 茎叶图：通常可用于统计和比较两组数据，其中茎是指中间的一列数，通常体现数据中除了末位数前面的其他数位，叶通常代表每个数据的末位数。并按末位数之前的数位进行分类排列，相同的数据需在茎叶图中体现多次

(四) 统计数据中的数字特征：

1、众数：一组数据中出现次数最多的数值，叫做众数

2、中位数：将一组数据从小到大排列，位于中间位置的数称为中位数，其中若数据的总数为奇数个，则为中间的数；若数据的总数为偶数个，则为中间两个数的平均值。

3、平均数：代表一组数据的平均水平，记为 \bar{x} ，设一组数据为： x_1, x_2, \dots, x_n ，则有：

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

4、方差：代表数据分布的分散程度，记为 s^2 ，设一组数据为： x_1, x_2, \dots, x_n ，其平均数为 \bar{x} ，

则有： $s^2 = \frac{1}{n} \left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right]$ ，其中 s^2 越小，说明数据越集中

5、标准差：也代表数据分布的分散程度，为方差的算术平方根

二、典型例题

例 1：某校高中部有三个年级，其中高三有学生 1000 人，现采用分层抽样法抽取一个容量为 185 的样本，已知在高一年级抽取了 75 人，高二年级抽取了 60 人，则高中部共有学生_____人。

思路：分层抽样即按比例抽样，由高一年级和高二年级的人数可得高三人数为 $185 - 75 - 60 = 50$ 人，所以抽样比为 $\frac{50}{1000} = \frac{1}{20}$ ，从而总人数为 $185 \div \frac{1}{20} = 3700$ 人

答案：3700

例 2：某企业三月中旬生产，A. B. C 三种产品共 3000 件，根据分层抽样的结果；企业统计