

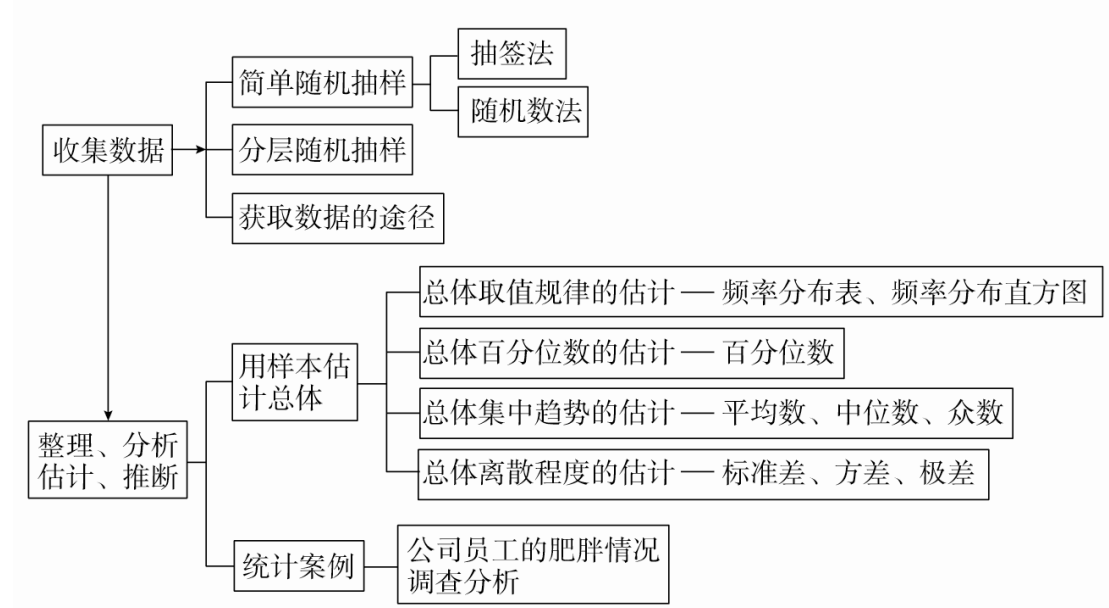
第九章 统计 全章复习

学习目标：

- 1. 整理归纳全章知识体系；
- 2. 进一步学习统计学用样本估计总体的核心思想，掌握用样本的数字特征估计总体的具体方法，理解各个统计量的意义。

学法指导：

一、本章知识系统



二、本章规律方法

1. 数理统计学的核心是如何根据样本的情况对总体的情况作出一种推断. 这里又包括两类问题：一类是如何从总体中抽取样本；另一类是如何根据样本的整理、计算和分析，对总体的情况作出判断. 抽样是手段、是先导，对总体的估计是目的、是结果. 学好这部分知识的关键是掌握正确的抽样方法，并会对总体进行科学的估计.

2. 随机抽样方法有简单随机抽样和比例分配的分层随机抽样这两种.

简单随机抽样和比例分配的分层随机抽样的共同特点是在抽样过程中每个个体被抽取的机会相等，体现了这些抽样方法的客观性和公平性. 其中简单随机抽样是最简单和最基本的抽样方法，在进行分层随机抽样时要用到简单随机抽样方法. 当总体中的个体数较少时，常采用简单随机抽样；当已知总体由差异明显的几部分组成时，常采用分层随机抽样.

实现简单随机抽样，常用抽签法和随机数法。

3. 对于所取不同数值较多或可以在实数区间内取值的总体，常用频率分布直方图来表示相应样本的频率分布。由于总体分布通常不易知道，我们往往是用样本的频率分布去估计总体分布。一般样本量越大，这种估计就越精确。

4. 统计的一个特征是通过部分的数据来推测全体数据的性质。由样本数据可以：

(1)列出频率分布表、画出频率分布直方图；

(2)求出样本数据的平均数、标准差、众数、中位数、百分位数等。

5. 注意掌握频率分布表的制作及画频率分布直方图的方法：

(1)求极差；

(2)决定组距与组数；

(3)将数据分组；

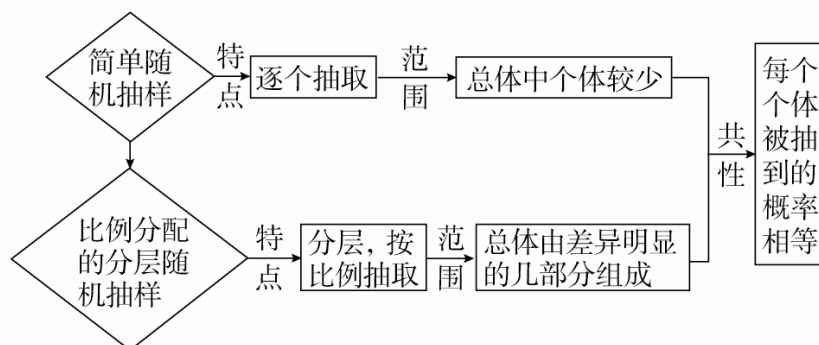
(4)列频率分布表；

(5)画频率分布直方图。

三、方法巩固提升

(一) 抽样方法的选取及应用

随机抽样方法有简单随机抽样和比例分配的分层随机抽样两种。



[例 1] 某校选修乒乓球课程的学生中，高一年级有 30 名，高二年级有 40 名，现用比例分配的分层随机抽样方法在这 70 名学生中抽取一个样本，已知在高一年级的学生中抽取了 6 名，则在高二年级的学生中应抽取的人数为()

A. 6

B. 8

C. 10

D. 12

解析 分层随机抽样的原理是按照各部分在总体中所占的比例抽取样本。设从高二年级抽取的学生数为 n ，

$$\text{则 } \frac{30}{40} = \frac{6}{n}, \text{ 解得 } n=8.$$

答案 B

(二) 利用样本的频率分布估计总体的取值规律

本章主要利用统计表、统计图分析估计总体的分布规律. 要熟练掌握绘制统计图表的方法, 明确图表中有关数据的意义是正确分析问题的关键, 从图形与图表中获取有关信息并加以整理.

[例 2] 为了让学生了解更多有关“一带一路”倡议的信息, 某中学举行了一次“丝绸之路知识竞赛”, 共有 800 名学生参加了这次竞赛. 为了解本次竞赛成绩情况, 从中抽取了部分学生的成绩(得分均为整数, 满分为 100 分)进行统计. 请你根据尚未完成的频率分布表, 解答下列问题:

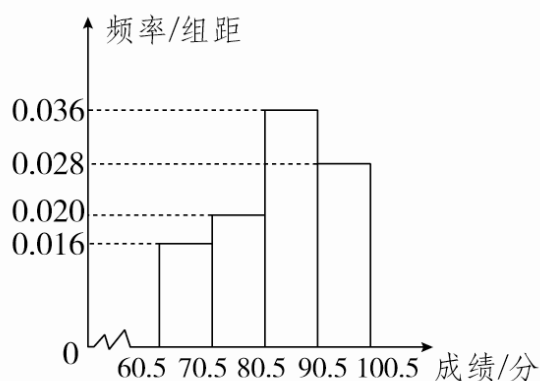
分组	频数	频率
60.5~70.5		0.16
70.5~80.5	10	
80.5~90.5	18	0.36
90.5~100.5		
合计		

(1) 填充频率分布表中的空格(将答案直接填在表格内), 并作出频率分布直方图;

(2) 若成绩在 85.5~95.5 分的学生为二等奖, 问参赛学生中获得二等奖的学生约有多少名?

解 (1) 频率分布表如下表所示, 频率分布直方图如图所示.

分组	频数	频率
60.5~70.5	8	0.16
70.5~80.5	10	0.20
80.5~90.5	18	0.36
90.5~100.5	14	0.28
合计	50	1



(2) 在被抽到的学生成绩中, 在 85.5~95.5 分的个数是 $9+7=16$, 占样本的比

例是 $\frac{16}{50}=0.32$ ，即获得二等奖的概率约为 32%，所以获得二等奖的学生约有 $800 \times 32\% = 256$ (名)。

(三) 利用样本的百分位数估计总体的百分位数

[例 3] 一家保险公司决定对推销员实行目标管理，即给推销员确定一个具体的销售目标。确定的销售目标是否合适，直接影响到公司的经济效益。如果目标定得过高，多数推销员完不成任务，会使推销员失去信心；如果目标定得太低，将不利于挖掘推销员的工作潜力。下面一组数据是部分推销员的月销售额(单位：千元)。

19. 58 16.11 16.45 20.45 20.24 21.66 22.45 18.22 12.34
19. 35 20.55 17.45 18.78 17.96 19.91 18.12 14.65 14.78
16. 78 18.78 18.29 18.51 17.86 19.58 19.21 18.55 16.34
15. 54 17.55 14.89 18.94 17.43 17.14 18.02 19.98 17.88
17. 32 19.35 15.45 19.58 13.45 21.34 14.00 18.42 23.00
17. 52 18.51 17.16 24.56 25.14

请根据这组样本数据提出使 65%的职工能够完成销售指标的建议。

解 将这 50 个样本数据按从小到大排序，可得

12. 34 13.45 14.00 14.65 14.78 14.89 15.45 15.54 16.11
16. 34 16.45 16.78 17.14 17.16 17.32 17.43 17.45 17.52
17. 55 17.86 17.88 17.96 18.02 18.12 18.22 18.29 18.42
18. 51 18.51 18.55 18.78 18.78 18.94 19.21 19.35 19.35
19. 58 19.58 19.58 19.91 19.98 20.24 20.45 20.55 21.34
21. 66 22.45 23.00 24.56 25.14

由 65%的职工能够完成销售指标，那么 35%的职工不能完成销售指标。

由 $50 \times (1 - 65\%) = 17.5$ 可知这组数据的 35%分位数为 17.52.故为使 65%的职工能够完成销售指标，该保险公司可将月销售额定为 17.52 千元。

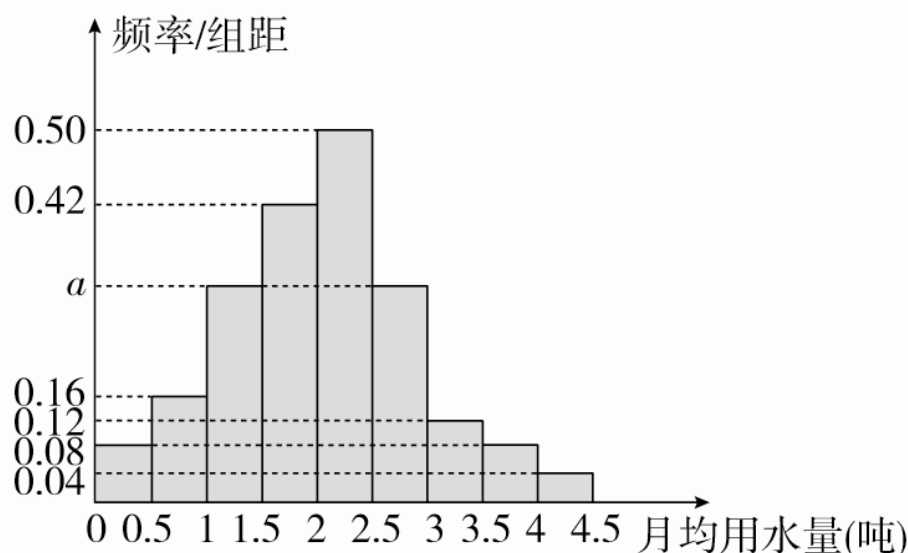
(四) 利用样本估计总体的集中趋势和离散程度

为了从整体上更好地把握总体的规律，我们还可以通过样本数据的众数、中位数、平均数和标准差等对总体作出估计。众数就是样本数据中出现次数最多的那个值；中位数就是把样本数据按照由小到大(或由大到小)的顺序排列，如果数据的个数是奇数，处于中间位置的数，如果数据的个数是偶数，中间两个数据的平均数；平均数就是所有样本数据的平均值，用 $\bar{x}$ 表示；标准差是反映样本数据

离散程度大小的最常用统计量，其计算公式是 $s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ 或者

$s = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}$ 有时也用标准差的平方 s^2 ——方差来代替标准差.

[例 4] 我国是世界上严重缺水的国家,某市为了制定合理的节水方案,对居民用水情况进行了调查,通过抽样,获得了某年 100 位居民每人的月均用水量(单位:吨),将数据按照 $[0,0.5)$, $[0.5,1)$, $\dots$, $[4,4.5]$ 分成 9 组,制成了如图所示的频率分布直方图.



(1)求直方图中 a 的值;

(2)设该市有 30 万居民,估计全市居民中月均用水量不低于 3 吨的人数,说明理由;

(3)估计居民月均用水量的中位数.

解 (1)由频率分布直方图,可知月均用水量在 $[0,0.5)$ 的频率为 $0.08 \times 0.5 = 0.04$.

同理,在 $[0.5,1)$, $[1.5,2)$, $[2,2.5)$, $[3,3.5)$, $[3.5,4)$, $[4,4.5]$ 组内的频率分别为 0.08, 0.21, 0.25, 0.06, 0.04, 0.02.

由 $1 - (0.04 + 0.08 + 0.21 + 0.25 + 0.06 + 0.04 + 0.02) = 0.5 \times a + 0.5 \times a$, 解得 $a = 0.30$.

(2)由(1), 100 位居民月均用水量不低于 3 吨的频率为 $0.06 + 0.04 + 0.02 = 0.12$, 由以上样本的频率分布,可以估计 30 万居民中月均用水量不低于 3 吨的人数为 $300000 \times 0.12 = 36000$.

(3)设中位数为 x 吨.

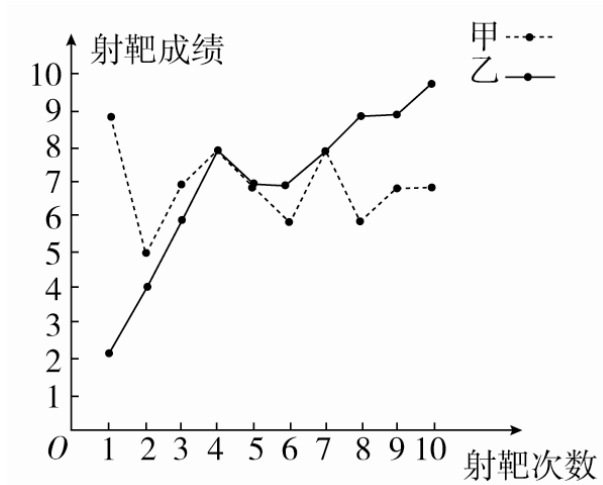
因为前 5 组的频率之和为 $0.04 + 0.08 + 0.15 + 0.21 + 0.25 = 0.73 > 0.5$,

而前 4 组的频率之和为 $0.04 + 0.08 + 0.15 + 0.21 = 0.48 < 0.5$, 所以 $2 \leq x < 2.5$.

由 $0.50 \times (x - 2) = 0.5 - 0.48$, 解得 $x = 2.04$.

故可估计居民月均用水量的中位数为 2.04 吨.

[例 5] 甲、乙两人在相同的条件下各射靶 10 次, 每次射靶成绩(单位: 环) 如下图所示.



(1)填写下表:

	平均数	方差	中位数	命中 9 环及以上
甲	7	1.2		1
乙		5.4		3

(2)请从四个不同的角度对这次测试进行分析:

- ①从平均数和方差结合分析偏离程度;
- ②从平均数和中位数结合分析谁的成绩好些;
- ③从平均数和命中 9 环以上的次数相结合看谁的成绩好些;
- ④从折线图上两人射击命中环数及走势分析谁更有潜力.

解 (1)乙的射靶环数依次为 2,4,6,8,7,7,8,9,9,10. 所以 $\bar{x}_乙 = \frac{1}{10} \times (2+4+6+8+7+7+8+9+9+10) = 7$; 乙的射靶环数从小到大排列为 2,4,6,7,7,8,8,9,9,10, 所以中位数是 $\frac{7+8}{2} = 7.5$; 甲的射靶环数从小到大排列为 5,6,6,7,7,7,7,8,8,9, 所以中位数为 7. 于是填充后的表格如下表所示:

	平均数	方差	中位数	命中 9 环及以上
甲	7	1.2	7	1
乙	7	5.4	7.5	3

(2)①甲、乙的平均数相同, 均为 7, 但 $s_甲^2 < s_乙^2$, 说明甲偏离平均数的程度小, 而乙偏离平均数的程度大. 甲比乙稳定些.

②甲、乙的平均水平相同, 而乙的中位数比甲大, 说明乙的射靶成绩比甲好些.

③甲、乙的平均水平相同，而乙命中 9 环以上(包含 9 环)的次数比甲多 2 次，可知乙的射靶成绩比甲好。

④从折线图上看，乙的成绩呈上升趋势，而甲的成绩在平均线上波动不大，说明乙的状态在提升，更有潜力。