9.2 用样本估计总体

FIG · B2-9-2

01 · 核心定理、公式与阶段边界

$$\text{矩形的高} = \frac{\text{频率}}{\text{组距}} \iff \text{矩形面积} = \text{高} \times \text{组距} = \text{频率} \iff \sum \text{矩形面积} = 1.0$$

核心本质：纵轴不是频率！必须比面积，组距不等时方能体现公平占比。

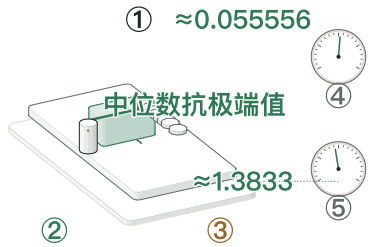
① 将 n 个数据从小到大排序；② 计算位置指标 $i = n \times p\%$；
③ **若 i 不是整数**：取第 $\lceil i \rceil$ 位数据（向上取整，如 $2.5 \to 3$）；
若 i 是整数：取第 i 位与第 $i+1$ 位两数的**算术平均值**。

直方图估计平均数: $\bar{x} = \sum(\text{组中值}_i \times \text{频率}_i)$

直方图估计中位数: 左右面积严格各占 0.5 的垂线位置

02 · 动手实验台

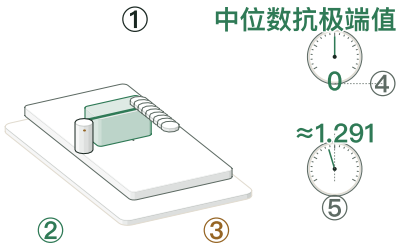
姿态 A · 基准状态



① 五数箱线图分拣台 · 样本 x₉=2.5，最大值 2.5
② 中位数 Q₂=0.00 (稳健不动!)
③ 均值 x̄=0.06, 标准差 s=1.38
④ 均值 x̄ · 量程[-1.5, 1.5]
⑤ 标准差 s · 量程[0, 3]

可调样本 x9 2.5 · 主体离散度 s 1.0

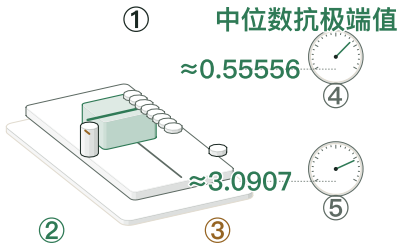
姿态 B · 下界/初始端



① 五数箱线图分拣台 · 样本 x₉=2，最大值 2
② 中位数 Q₂=0.00 (稳健不动!)
③ 均值 x̄=0.00, 标准差 s=1.29
④ 均值 x̄ · 量程[-1.5, 1.5]
⑤ 标准差 s · 量程[0, 3]

可调样本 x9 2.0 · 主体离散度 s 1.0

姿态 C · 上界/满量程端



① 五数箱线图分拣台 · 样本 x₉=8，最大值 8
② 中位数 Q₂=0.00 (稳健不动!)
③ 均值 x̄=0.56, 标准差 s=3.09
④ 均值 x̄ · 量程[-1.5, 1.5]
⑤ 标准差 s · 自动量程[0, 4]

可调样本 x9 8.0 · 主体离散度 s 1.5

03 · 诊断与易错辨析自测（先独立作答，再核对答案页）

1. 样本量是 200，频率分布直方图中 $[20, 30)$ 这一组矩形的高是 0.025。这一组的频数是?

☐ A 5　☐ B 25　☐ C 50　☐ D 0.25

2. 12 名同学一分钟跳绳（次），已排好序：120，125，130，132，135，138，140，142，145，150，155，160。第 70 百分位数是?

☐ A 142　☐ B 143.5　☐ C 145　☐ D 150

3. 一张直方图的分组不等宽：$[0, 10)$ 这一组的高是 0.02，$[10, 30)$ 这一组的高也是 0.02。哪一组的频率大?

☐ A一样大，因为一样高　☐ B $[0, 10)$ 大　☐ C $[10, 30)$ 大

4. 按本书的定义，数据 $1, 3, 5, 7$ 的方差是?

☐ A 5　☐ B $\frac{20}{3}$　☐ C 20　☐ D $\sqrt{5}$

参考答案与防坑要点：

(1) [C]

C。频率 $= 0.025 \times 10 = 0.25$，频数 $= 200 \times 0.25 = 50$。选 A 是把高 0.025 当频率直接乘了 200；选 D 只算到频率，没乘样本量。

(2) [C]

C。$i = 12 \times 70\% = 8.4$，不是整数，往上取整，取第 9 位：145。选 A 是往下取了第 8 位；选 B 是把“取两位平均”用在了 i 不是整数的时候。

(3) [C]

C。比的是面积：$0.02 \times 10 = 0.2$，$0.02 \times 20 = 0.4$，后者是前者的 2 倍。纵轴是“频率÷组距”，组距不同时只看高度会看错。

(4) [A]

A。$\bar{x} = 4$，差的平方 $9, 1, 1, 9$，和 20，除以 $n = 4$ 得 5。B 是除以 $n - 1 = 3$，本书不用这个分母；C 忘了除；D 是标准差，不是方差。