

1 数理统计概念

统计学的核心：用样本推断总体。

1.1 样本与总体

研究对象的全体称为总体，从总体中抽取的一部分个体叫做样本。

1.2 统计量

统计量是样本的一个函数 $T(X_1, X_2, \dots, X_n)$

常用的统计量有：

1. 样本均值： $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
2. 样本方差： $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
3. k 阶原点矩： $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$
4. k 阶中心矩： $M_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$

5. 顺序统计量: $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$

2 描述性统计

数据集中趋势: 平均数, 中位数, 频数, 众数, 百分位数

数据离散趋势: 方差, 标准差, 极差, 四分位差

分布特征:

1. 分布函数: $F(x) = P\{X \leq x\}, x \in (-\infty, +\infty)$

2. 概率密度函数: $f(x) = F'(x)$

偏度与丰度:

1. 偏度 (skewness) 也称为偏态, 刻画数据的对称性。记为 bs , $bs < 0$ 左偏, $bs > 0$ 有偏。

2. 峰度 (kurtosis) 是分布曲线在均值处高低的特征数。刻画的是分布函数的集中和分散程度。