

第六章、统计学中的基本概念

§6.1 引言

- 数理统计学: 收集、整理、分析带随机性的**数据**(data).
- 例如, 学校希望了解两名学生的学习能力有没有差别, 他们的成绩表如下.

学生	语文	数学	英语	物化
甲	80	70	60	85
乙	70	90	65	80

- 重点: **方法!**

§6.2 若干基本概念

- 总体指所考查的对象的总和, 归结为随机变量 X .
- 总体分布(族)指 F_θ , $\theta \in \Theta$, 其中 θ 为参数, Θ 为参数空间. 可表现为分布密度、分布列、分布函数.
- 总体模型指 $X \sim F_\theta$, $\theta \in \Theta$.
- (简单随机)样本指 $\vec{X} = (X_1, \dots, X_n)$, 其中 X_1, \dots, X_n 独立同分布, $X_1 \sim F_\theta$.
- \vec{X} 的分布(对应的概率)记为 P_θ .
- 统计模型指 \vec{X} 与 P_θ , $\theta \in \Theta$.
- 样本量指 n . 样本值指 \vec{X} 的具体取值 $\vec{x} = (x_1, \dots, x_n)$.
- \vec{X} 着重方法/理论分析; \vec{x} 着重应用/计算.

例2.1. 对某待估量 a 重复独立测量, 得到测量值 x_1, \dots, x_n .

- 总体分布: $X = a + e$, 其中 $e \sim N(0, \sigma^2)$. 即, $X \sim N(a, \sigma^2)$.
- 参数: $\theta = (a, \sigma^2)$. 其中, σ^2 不是所关心的, 称为讨厌参数.
- P_θ : \vec{X} 的联合密度为

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i - a)^2}{2\sigma^2}\right\}.$$

- 研究对象: θ 或 $g(\theta)$. 例, $g(a, \sigma^2) = a$.
- 统计量指 \vec{X} 的(实值或向量值)函数 $T(\vec{X})$ (定义2.3).
例, 用 $T(\vec{X}) := \frac{1}{n}(X_1 + \dots + X_n)$ 来估计 a .

§6.3 若干统计问题

- 参数估计: 给出 θ 或 $g(\theta)$ 的估计量 $\hat{\theta} = T(\vec{X})$.
- 例3.1. $X \sim N(a, \sigma^2)$, σ^2 为讨厌参数.

$$\hat{a} = T(X) = \frac{1}{n}(X_1 + \cdots + X_n) \sim N(a, \frac{1}{n}\sigma^2).$$

- 假设检验: 对假设 H_0 回答“是”或“否”.
例, 规定不合格率不能超过3%. 现有200件产品, 从中任意抽取10件, 发现2件不合格. 问: 是否可以出厂?
- 线性回归: 研究变量 Y 对 x 的线性依赖关系,

$$Y = b_0 + b_1x + e, \quad e \sim N(0, \sigma^2).$$