



专注于商业智能BI和大数据的垂直社区平台

假设检验的基本概念（一）

Allen

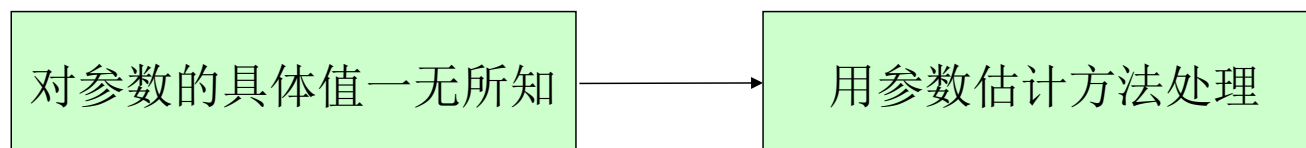
www.hellobi.com

课程目录

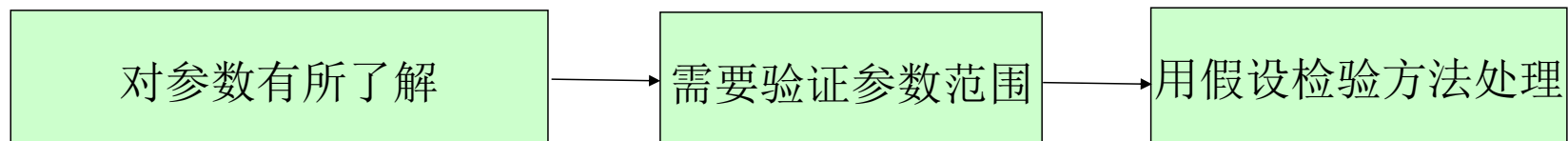
- 统计推断的两类问题
- 引例
- 假设检验的理论依据
- 引例所引出的基本概念
- 小结

统计推断的两类问题

- 问题一：参数估计



- 问题二：假设检验



引例

- 设某厂生成一种灯泡，其寿命服从正态分布 $X \sim N(\mu, 40000)$ ，过去较长时间来看，灯泡的平均寿命为1500小时。但是现在采用新生产线后，从生产的灯泡中随机抽取了25只，测得平均寿命为1700小时。问采用新生产线后，灯泡的寿命是否显著提高？

引例

- 1.上述问题为检验假设问题而非参数估计问题
- 2.命题“灯泡寿命是否大于1500小时”是否正确涉及到如下两个参数集

合 $H_0: \mu = 1500$ $H_1: \mu > 1500$

将这两个非空参数集合称为**统计假设**，简称假设

引例

- 3.上述问题第一个统计假设 $\mu = 1500$ 表示采用新产品线后灯泡寿命没有显著增加；第二个统计假设 $\mu > 1500$ 表示采用新产品线后灯泡寿命有显著增加
- 4.将第一个假设称为**原假设**，用符号 $H_0: \mu = 1500$ 表示
将第二个假设称为**备择假设**，用符号 $H_1: \mu > 1500$ 表示

在假设检验中，常把一个被检验的假设称为原假设，或者是希望从子样观测值得到对某一陈述的强有力支持，就把这一陈述的否则作为原假设，另一个为备择假设

假设检验的理论依据

- 假设检验之所以可行，其理论背景为实际推断原理，即“小概率原理”
- 通常借助于直观分析和理论分析相结合的做法，其基本原理就是人们在实际问题中经常采用的所谓实际推断原理：“一个小概率事件在一次试验中几乎是不可能发生的”

引例所引出的基本概念

- **参数假设**：母体的分布类型为已知，仅有一个或几个参数为未知，只对未知参数作出假设，这种仅涉及到母体分布的未知参数的统计假设称为参数假设

引例中对灯泡寿命 μ 进行的统计假设

- **非参数假设**：只对未知分布函数的类型或者它的某些特征提出某种假设，称这种统计假设为非参数假设

若灯泡寿命所服从的分布未知，对寿命是否服从正态分布作出假设

引例所引出的基本概念

- **简单统计假设/简单假设**：如果一个统计假设完全确定母体的分布，则称这种假设为简单统计假设

引例中 $H_0: \mu = 1500$ 完全确定母体分布 $N(1500, 40000)$ ，所以是简单假设

- **复杂统计假设/复杂假设**：如果一个统计假设不能完全确定母体的分布，则称这种假设为复杂统计假设

引例中 $H_1: \mu > 1500$ 不能完全确定母体分布，所以是复杂假设

引例所引出的基本概念

- 假设检验问题的一般提法：在给定备择假设 H_1 下对原假设 H_0 作出判断，具体判断是拒绝原假设 H_0 还是接受原假设 H_0
- **检验法则**：在 H_0 对 H_1 的检验问题中要作出某种判断，必须要从子样出发，指定一个法则，一旦子样观察值确定后，利用构造的法则作出判断：拒绝 H_0 还是拒绝 H_1 ，这种法则就称为 H_0 对 H_1 的一个检验法则

检验法则是什么呢？

引例所引出的基本概念

- 假设的法则本质上是子样的一个统计量，称为**检验统计量**，最后根据观察值在统计量上的值来决定是否拒绝原假设
- **拒绝域/临界域**：将子样空间划分成两个互不相交的子集 C 和 C^* ，使得子样的观察值 $(x_1, x_2, \dots, x_n) \in C$ 时拒绝原假设 H_0 ， $(x_1, x_2, \dots, x_n) \in C^*$ 时，接收原假设 H_0 ，这样的划分就构成一个准则，我们称这个子样空间的子集 C 为检验的临界域/拒绝域，拒绝域的边界点成为临界点

小结

- 统计推断的两类问题
- 引例
- 假设检验的理论依据
- 引例所引出的基本概念
- 小结