概率论与数理统计

第八章

韩潇

xhan011@ustc.edu.cn

第八章： 假设检验

大纲：

- 基本概念
- 正态总体参数检验
- 比例p的检验
- 似然比检验
- P值

8.1 基本概念

- 通过下面例子来给出假设检验问题提法的形式

例 8.1 按质量标准要求, 合格的棉布单位长度中的瑕疵点不超过 2 个. 从一批棉布中随机抽检 10 包棉布, 得到棉布单位长度平均瑕疵点为 3 个. 据此数据回答这批棉布质量是否符合质量要求.

➢ 对统计总体（即总体分布）的性质所作的假设称为统计假设. 使用样本对所作出的统计假设进行检查的方法和过程称为假设检验.

- 总体分布的类型是已知的，要检验的假设是有关总体参数的某个取值范围，就称为参数假设检验问题
- 如果总体分布类型完全未知, 我们称之为非参数假设检验问题

8.1原假设和备择假设

- 在统计学中，我们把关于总体分布的某个特征的假设命题称为一个“假设”或“统计假设”，例如假设总体分布为正态分布，假设总体分布为二项分布等

- 称之为“假设”就是这个命题是否成立还需要通过样本来检验

- 一般把认为是正确的命题称为“原假设”(Null hypothesis)，记为H_0

- 当原假设不成立时你准备接受什么结论？这个假设称为“备择假设”(alternative hypothesis)，记为H_1或H_a，就是拒绝原假设后可供选择的假设

8.1原假设和备择假设

例 8.2 在一条食盐生产线上, 已经调好每袋食盐的重量是 350 克, 由于长期生产, 机器包装的每袋食盐重量可能发生变化. 我们随机从生产线上抽取 12 袋进行检查, 要从中得出生产线工作是否正常. 在这个问题中, 由于是早已调好每袋 350 克, 所以正常生产时每袋重量可以假定为正态分布 $N(350,\sigma^2)$, 其中 σ^2 可以设定好, 也可能会随时间而变化. 我们的重点是看均值有无变化, 所以原假设就是 $H_0: \mu = 350$(克). 如果根据样本发现原假设 H_0 不成立, 每袋食盐重量可能在 350 克上下波动, 因此备择假设为 $H_1: \mu \neq 350$(克). 另一种可能这个企业是良心企业, 每袋只会倾向于多一点, 不会少于 350 克, 这时的备择假设就是每袋是否太多了, 即 $H_1: \mu > 350$(克). 当然也有另一种可能, 即只少不多. 这时检验的备择假设就是 $H_1: \mu < 350$(克).

8.1原假设和备择假设

一般地，如果记Θ_0和Θ_1是参数空间$\Theta \subseteq R^{\mathrm{k}}$的两个不交非空子集，一个统计假设常表示为

$$H_0: \theta \in \Theta_0 \leftrightarrow H_1: \theta \in \Theta_1 \tag{8.1}$$

其中$\Theta_0 \cup \Theta_1 \subseteq \Theta$

8.1简单假设和复合假设

不论是原假设还是备择假设，其中的假设只有一个参数值，就称为简单假设，否则称为复合假设. 如果记感兴趣的参数为$\theta \in \Theta \subseteq R^{k}$, 则常见关于$\theta$的假设形式有

$$(1)\ H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta = \theta_1 \tag{8.2}$$

$$(2)\ H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0 \tag{8.3}$$

$$(3)\ H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0 \text{ 或 } H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta > \theta_0 \tag{8.4}$$

$$(4)\ H_0 : \theta \geq \theta_0 \leftrightarrow H_1 : \theta < \theta_0 \text{ 或 } H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta < \theta_0 \tag{8.5}$$

其中假设 (1) 也称为两点假设, (2) 称为双边假设（two-sided hypothesis）, (3) 和 (4) 称为单边假设（one-sided hypothesis）.

8.1检验统计量、接受域、拒绝域和临界值

- 在检验一个假设时用到的统计量称为检验统计量

- 使原假设得到接受的样本所在区域，称为该检验的接受域，而使原假设被拒绝的样本所在区域，称为拒绝域(或否定域)

- 记样本$\boldsymbol{X}=(X_1,\ldots,X_n)$, 则接受域$\boldsymbol{A}=\{(X_1,\ldots,X_n)\in A\}$和拒绝域$\boldsymbol{D}=\{(X_1,\ldots,X_n)\in D\}$都是 R^n中的一个区域

- 常见的假设检验中接受域和拒绝域通常可以简化为检验统计量$T(\boldsymbol{X})$所处的区域

8.1检验统计量、接受域、拒绝域和临界值

- 例：棉布质量检验接受域$\boldsymbol{A}=\{\bar{\boldsymbol{X}}\leq C\}$，而拒绝域为$\boldsymbol{D}=\{\bar{\boldsymbol{X}}>C\}$. 由于$\boldsymbol{A},\boldsymbol{D}$互补, 知道其中一个就能知道另一个. 因此，指出其中之一即可
- 常数C处于一种特殊的位置， 检验统计量$\bar{\boldsymbol{X}}$越过C， 结论就由接受变为拒绝，这个C称为临界值． 也可能有不止一个临界值.
- 这些决策也可以用函数

$$\Psi(X_1,\ldots,X_n)=\begin{cases}1, & (X_1,\ldots,X_n)\in D\\ 0, & (X_1,\ldots,X_n)\in A\end{cases}$$

来表示，称Ψ为对H_0和H_1的一个检验函数（法则）.

8.1功效函数

- 同一个原假设， 可以有不同的检验方法， 哪一种更好一点?为此，我们引入功效函数(power function)的概念

定义 8.1 功效函数

设总体为 $F(x,\theta)$, 其中 $\theta=(\theta_1,\ldots,\theta_k)\in\Theta\subseteq\mathbb{R}^k$ 为参数, H_0 是关于参数 θ 的一个原假设, 设 Ψ 是根据样本 $(X_1,\ldots,X_n)$ 对假设(8.1)所作的一个检验, 则称

$$\beta_\Psi(\theta)=\mathbb{P}_\theta(\text{在检验}\Psi\text{下}H_0\text{被否定}) \tag{8.7}$$

为检验 Ψ 的功效函数. ♣

- 功效函数是参数θ的函数,
- 原假设成立时,我们不希望拒绝它,而原假设不成立时,我们希望拒绝它

8.1功效函数

定义 8.2 检验水平

设 Ψ 是假设(8.1)的一个检验, $\beta_\Psi(\theta)$ 为其功效函数, α 为常数, $0 \le \alpha \le 1$. 如果

$$\beta_\Psi(\theta) \le \alpha, \quad 对任何\theta \in H_0, \tag{8.8}$$

则称 Ψ 为 H_0 的一个水平 α 的检验, 或者说, 检验 Ψ 的水平为 α(或检验 Ψ 有水平 α).

♣

若α为Ψ的水平而$\alpha_1 > \alpha$，则α_1也是检验的水平，即水平不唯一. 为了克服这点不方便之处，通常只要有可能就选择最小可能的水平作为检验的水平.

8.1两类错误

由于样本有随机性，所以 Ψ必犯以下两类错误之一

1. 当 H_0 正确时, 但是检验法则 Ψ 拒绝了 H_0, 这称为检验 Ψ 犯了第一类错误, 也称“弃真错误”. 其概率记为$\alpha_{1\Psi}(\theta)$

2. 当 H_0 错误时, 但是检验法则 Ψ 没有拒绝 H_0，这称为检验 Ψ 犯了第二类错误，也称“存伪错误”. 其概率记为$\alpha_{2\Psi}(\theta)$

根据接受域 A 和拒绝域 D 的定义, 我们只能犯两种错误之一. 这两种错误与功效函数有如下关系:

$$\alpha_{1\Psi}(\theta)=\begin{cases}\beta_{\Psi}(\theta), & 当\theta\in H_0,\\ 0, & 当\theta\in H_1,\end{cases} \tag{8.9}$$

8.1两类错误

$$\alpha_{2\Psi}(\theta) = \begin{cases} 0, & 当\theta \in H_0, \\ 1 - \beta_\Psi(\theta), & 当\theta \in H_1. \end{cases} \tag{8.10}$$

我们希望犯两类错误的概率都尽量小,但这是不可能的

- 奈曼提出先保证犯第一类错误的概率不超过某个给定的很小的数α，在此基础上使犯第二类错误的概率尽量小. 如果仅仅考虑控制第一类错误的概率，而不涉及犯第二类错误概率所得到的检验，我们称为显著性检验(significance test).

- 定义8.2中检验的水平就是用这个检验时犯第一类错误的概率，我们把检验中犯第一类错误的概率控制在$(0, \alpha]$内，α也称为显著性水平

8.1两类错误

例 8.3 某饮料厂在自动流水线上罐装饮料. 在正常生产情况下, 每瓶饮料的容量 (单位: 毫升) X 服从正态分布 $N(500, 10^2)$ (由以往的经验得知). 经过一段时间之后, 有人觉得每瓶饮料的平均容量减小到 490, 于是抽取了 9 瓶样品, 称得它们的平均值为 $\bar{x} = 492$ 毫升. 能否在显著性水平 0.05 下认为饮料的平均容量确实减少到 490 毫升?

8.1显著性检验方法的一般步骤

第 1 步: 求出未知参数 θ 的一个较优的点估计 $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$, 如最大似然估计.

第 2 步: 以 $\hat{\theta}$ 为基础, 寻找一个检验统计量

$$T = T(\hat{\theta}, \theta_0)$$

使得当 $\theta = \theta_0$ 时, T 的分布已知 (如 $N(0,1), t_n, F_{m,n}$) , 从而容易通过查表或计算得到这个分布的分位数, 用以作为检验的临界值.

第 3 步: 以检验统计量 T 为基础, 根据备择假设 H_1 的实际意义, 寻找适当形状的拒绝域 (它是关于 T 的一个或两个不等式, 其中包含一个或两个临界值).

第 4 步: 当原假设成立时, 犯第 I 类错误的概率小于或等于给定的显著性水平 α, 这给出一个关于临界值的方程, 解出临界值, 它 (们) 等于 T 在 $\theta = \theta_0$ 的分布的分位数, 这样即确定了检验的拒绝域.

8.1显著性检验方法的一般步骤

第 5 步: 如果给出样本值, 则可算出检验统计量的值, 如果落在拒绝域中则可拒绝原假设, 否则不能拒绝原假设.

第 6 步: 根据具体问题和给定的显著性水平 α 解释拒绝原假设或不能拒绝原假设.

Thank you!