第七讲：假设检验

7.1 基本概念和问题的提法

- (统计) 假设: 在数理统计中, 关于总体分布的概率性质的假定. 例如假设正态总体, 二项总体等, 或者二项总体中成功概率 $p \le 0.5$ 等等.
- (统计) 检验: 使用样本对所作出的假设进行检查的方法和过程.

7.1.1 基本概念

假设检验问题就是研究如何根据抽样后获得的样本来检查抽样前所作假设是否合理.

首先, 由一个例子引出一些基本概念.

↑Example

某厂产品出厂检验规定: 每批产品次品率 p 不超过 4% 才能出厂。现从某批产品 10000 件中任意抽查 12 件发现 4 件次品, 问该批产品能否出厂? 若抽得结果是 1 件次品呢?

↓Example

解： 若以 p 表示此批产品的次品率，则问该批产品能否出厂等价于即要检验次品率 p 是否不超过 4%。我们假设 "$p \le 4\%$"，并记 Y 为 12 件中的次品数，由于总产品数很大，故可以认为 $Y \sim B(12, p)$，此时当 $p \le 0.04$ 时，

$$P(Y=4) = \binom{12}{4} p^4 q^8 < \binom{12}{4} 0.04^4 0.96^8 = 0.000914$$

这是一个小概率事件，即当 $p \le 0.04$ 时，12 件产品中有 4 件是次品的概率不到 1/1000，这样的事件在一次试验中几乎是不可能发生的，但确实发生了 (我们观察到了 4 件次品), 因此更倾向于怀疑假设 "$p \le 0.04$" 的正确性，即认为它不成立。而由于

$$P(Y=1) \le \binom{12}{1} 0.04^1 0.96^{12} = 0.306$$

即此时当假设 "$p \le 0.04$" 成立时，"12 个产品中有一个次品" 这一事件的概率最大为 0.306，这个事件不是小概率事件。因此我们没有足够的证据支持原假设不成立这一说法。

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. 试问此断言是否正确? 即问平均每瓶饮料的容量仍是 500 毫升还是变成 490 毫升? 假定标准差 10 毫升不变.

↑Example

↓Example

在这个问题中,

统计假设: 罐装饮料容量 $X \sim N(\mu, 10^2)$.

问题: 根据样本来在 "$\mu = 500$" 和 "$\mu = 490$" 之间作判断.

数理统计中, 把它们看成两个假设. 习惯上, 称前者为**原假设或零假设**, 记作 H_0; 后者称为**备择假设或对立假设**, 记作 H_1 或 H_a. 所谓检验

$$H_0 : \mu = 500 \leftrightarrow H_1 : \mu = 490.$$

就是要根据样本判断究竟是 “H_0 成立” 还是 “H_1 成立”. 断言 “H_0 成立” 称为**不能拒绝** H_0; 断言 “H_1 成立” 称为**拒绝** H_0.

下面讨论如何检验上述假设, 即给定一个接受或者拒绝零假设的准则. 设从总体中抽取一个样本 $X_1, \cdots, X_n$, 我们可以用极大似然估计 $T = \bar{X}$ (称之为**检验统计量**) 来估计 μ. 由于该估计值接近 μ (尤其是当样本量较大时), 故当 T 的绝对值小的时候有利于 H_1 而不利于 H_0, 此时应该拒绝 H_0. 我们可以事先取定一个常数 τ, 称之为**临界值**, 当 T 的取值小于该临界值时拒绝 H_0, 即样本满足

$$W = \{\bar{X} < \tau\}$$

中时拒绝 H_0, 称 W 为**拒绝域**. 即样本的取值落在拒绝域中, 就拒绝 H_0, 否则不能拒绝之. 一个拒绝域就对应于一个检验方法. 现在的问题是 τ 应该取多大? 这涉及到**两类错误**.

决策 \ 事实	H_0 成立	H_1 成立
不拒绝 H_0	不犯错	第 II 类错误
拒绝 H_0	第 I 类错误	不犯错

- **称 "实际上 H_0 成立但是它被拒绝" 这个错误**为**第 I 类错误 (弃真)**
- **"实际上 H_0 不成立但是它没有被拒绝" 这样一类错误**为 **第 II 类错误 (存伪)**.

而由于我们的方法是基于观测数据, 而观测数据是带有随机误差的, 故难免在做出决策的时候犯错, 我们能做的是控制犯错的概率.

一个理想的检验应该使这两类错误的概率都小, 但是在实际问题中不可能使这两类错误一致地小: 要让犯第 I 类错误的概率小, 应该让 τ 小, 而要让犯第 II 类错误的概率小, 则 τ 不能太小. 解决这个矛盾的一个方法是**在控制 I 类错误的基础上, 尽量少犯第 II 类错误**

(在下一小节中我们讨论如何设定假设时会提到, 应该将受保护对象设为零假设, 故犯第 I 类错误的严重性更大, 因此必须尽量避免犯第 I 类错误). 因此, **这种在只限制第一类错误的原则下的检验方法, 就称为 "显著性检验"** (Significance Test)。

具体地, 给定一个允许的犯第一类错误概率的最大值 α, 选取 τ 使得

$$P_{H_0}(T < \tau) \le \alpha$$

连续场合下, 等号可以达到. 这样的 τ 可以通过 T 在 H_0 下的分布及上式条件下求得.

称 α 为**显著性水平**. 通常将 α 取为 0.1, 0.05, 0.01 等较小的数, 具体取值视实际需要而定, 有时候要求 α 很小, 比如在涉及到数十万个基因标记的基因关联分析中, 单个位点检验的 α 一般是 10^{-7} 这样的量级.

现在将问题一般化:

1. 根据问题, 提出假设检验问题

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1.$$

   其中 H_0 为**零假设**或**原假设**, 而 H_1 为**对立假设**或**备择假设**.

2. 根据参数的估计方法构造一个适当的**检验统计量** $T = T(X_1, \cdots, X_n)$, 其中 $X_1, \cdots, X_n$ 为从总体中抽得的一个样本.

3. 根据对立假设的形状构造一个检验的**拒绝域** $W = \{T(X_1, \cdots, X_n) \in A\}$, 其中 A 为一个集合, 通常是一个区间. 比如拒绝域可取为 $\{T(X_1, \cdots, X_n) > \tau\}$, 则称 τ 为 **临界值**.

4. 对任意的 $\theta \in \Theta_0$, 犯第 I 类错误的概率 $P_\theta(T(X_1, \cdots, X_n) \in A)$ 小于或等于某个指定正的常数 α), 则称 α 为**显著性水平**.

5. 结合 T 在 H_0 下的分布, 定出 A.

显然显著性水平不是唯一的, 事实上, 如果 α 是一个显著性水平, 则任意大于 α 的数都是显著性水平. 实际中通常采用显著性水平最小的那一个.

称 $\beta(\theta) = P_\theta$（H_0 被拒绝）为检验的**功效函数**. Definition

如果检验的显著性水平为 α, 则当 $\theta \in \Theta_0$ 时, $\beta(\theta) \leq \alpha$. 而当 $\theta \in \Theta_1$ 时, 我们希望功效值越大越好 (这样犯第 II 类错误的概率 $1 - \beta(\theta)$ 就越小), 所以功效可以作为评价一个检验优劣的准则.

显著性检验方法由于只控制第一类错误, 因此保护了原假设不被轻易拒绝, 从而原假设和对立假设地位不相同!

7.1.2 原假设的提法

在有时候需要自己判断如何提假设检验问题. 在建立原假设时有两个原则。

原则一：将受保护的对象置为零假设. 将已经存在的事实, 或者错误拒绝会带来很大后果的事情作为原假设. 例如, 司法上的无罪推断. 这样做大大地有利于保护公民的利益. 又若对新药的批准, 显然使用药品的病人是应该受保护的对象, 这时应该设定一个有利于病人的命题作为零假设, 这个命题就是 “新药不比安慰剂效果好”, 以尽量避免病人用无效甚至有副作用的新药. 将检验的显著性水平 α 设定得较小, 以保证零假设不被轻易推翻.

在实际问题中, 如果根据某个合理的检验方法发现零假设被推翻, 则有充分的理由认为零假设不成立而对立假设成立, 这是因为万一零假设成立而被误据的概率不会超过 α; 另一方面, 如果发现零假设未被拒绝, 并不表明有充分理由接受零假设, 而是因为零假设被保护得较严密以至于未被拒绝.

原则二：如果你希望"证明"某个命题，就取相反结论或者其中一部分作为零假设 (类似于反证法). 这种提法往往是在两个假设命题中不太清楚哪个应受保护，此时可以借用司法制度里的"谁主张，谁举证"，即若想用统计方法向人"证明"一个命题，则将那个命题置为对立假设. 注意这里的证明不是数学上的严格证明，而是允许犯错的一种统计推断方法. 用统计方法证明一个命题不是一件容易的事情，所以如果没有足够把握，人们应该避免用统计方法去证明一个命题.

原则三：假设检验的"拒绝零假设"结果比"不能拒绝零假设"更有保证. 当我们没有任何知识来选择零假设时候，提一个零假设使得使用某个检验方法得到的结果为"拒绝零假设". 因为零假设不被轻易拒绝，其犯错的概率不超过 α，因此得到"拒绝零假设"的结论比得到"不能拒绝零假设"的结论更有保证.

7.1.3 检验统计量的选取及假设检验的步骤

通过解答前例来说明假设检验的步骤.

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?

↑Example

↓Example

解: 基于统计量 $\bar{X}$, 我们采用 "标准化" 过的检验统计量 (减均值再除以标准差)

$$T_1 = \frac{\sqrt{n}(\bar{X} - 500)}{10}$$

以使该统计量服从标准正态分布, 检验的拒绝域仍取形如 $\{T_1 < \tau_1\}$, 我们控制犯第 I 类错误的概率等于 α, 即

$$P(T_1 < \tau_1 | \theta = 500) = \alpha.$$

由于 $\theta = 500$ 时 T_1 服从标准正态分布, 易知上面关于 τ_1 的方程的解为 $\tau_1 = -u_\alpha$, 其中 u_c 等于标准正态分布的上 c 分位数, 即检验的拒绝域为

$$\{T_1 < -u_\alpha\}.$$

现在取显著性水平为 0.05, 则临界值 $u_{0.05} \approx 1.645$. 另一方面, 样本均值 $\bar{x} = 492$, 样本量 $n = 9$, 故检验统计量 T_1 的观测值等于 -2.4, 小于临界值 1.645, 即样本落在拒绝域中, 从而可以在显著性水平 0.05 下拒绝零假设, 认为饮料的平均容量确实减少为 490 毫升.

下面列举几种常见的假设检验问题:

(1) $H_0: \ \theta = \theta_0 \leftrightarrow H_1: \theta = \theta_1$;

(2) $H_0: \ \theta = \theta_0 \leftrightarrow H_1: \ \theta \neq \theta_0$;

(3) $H_0: \ \theta = \theta_0 \leftrightarrow H_1: \ \theta > \theta_0$ 或者 $H_0: \ \theta \leq \theta_0 \leftrightarrow H_1: \ \theta > \theta_0$

(4) $H_0: \ \theta = \theta_0 \leftrightarrow H_1: \ \theta < \theta_0$ 或者 $H_0: \ \theta \geq \theta_0 \leftrightarrow H_1: \ \theta < \theta_0$

称 (1) 为**简单假设**, (2) 为**双侧假设**因为对立假设是双侧的, (3) 和 (4) 为**单侧假设**因为对立假设是单侧的. 这里强调对立假设的原因是检验方法 (对应于一个拒绝域) 只跟对立假设有关.

对上述这些假设, 我们总结一下显著性检验的一般步骤, 设定显著性水平为 α.

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

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

$$T = t(X_1, \cdots, X_n)$$

且使得当 $\theta = \theta_0$ 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, 根据对立假设 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: 当零假设成立时, 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 的分位数, 这样即确定了检验的拒绝域.

第 5 步: 如果给出样本观测值, 则可算出检验统计量的样本观测值, 如落在拒绝域中则可拒绝零假设, 否则不能.