

目 录

第七章 假设检验	1
§7.1 基本概念和问题的提法	1
§7.1.1 零假设, 对立假设, 两类错误, 拒绝域, 显著性水平, 功效	1
§7.1.2 原假设的提法	4
§7.1.3 检验统计量的选取及假设检验的步骤	4

第七章 假设检验

教学目的:

- 1) 理解假设检验的一些基本概念: 零假设、对立假设、两类错误、拒绝域、显著性水平、功效.
- 2) 学会将实际问题转化成假设检验问题来处理.
- 3) 一样本和两样本正态总体均值和方差的假设检验.
- 4) 0-1 分布参数的假设检验.
- 5) 拟合优度检验、列联表的独立性和齐一性检验.

§7.1 基本概念和问题的提法

§7.1.1 零假设, 对立假设, 两类错误, 拒绝域, 显著性水平, 功效

在参数估计问题中, 常常在抽样前先对未知总体作一些假定. 例如假定总体 X 服从正态分布, 假定某个正态总体的方差为一个已知值等等. 在数理统计中, 关于总体分布的概率性质的假定称为(统计) 假设. 抽样前所作出的假设是否与实际符合, 可以用样本所提供的信息来检查, 检查的方法与过程称为(统计) 检验. 假设检验问题就是研究如何根据抽样后获得的样本来检验抽样前所作出的假设. 首先, 由一个例子引出一些基本概念.

例 7.1.1. 某厂产品出厂检验规定: 某批产品次品率 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?

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“ $p \leq 4\%$ ”, 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 $p \leq 0.04$ 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 $p \leq 0.04$ 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“ $p \leq 0.04$ ”的正确性, 即认为它不成立。而由于

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

即此时当假设“ $p \leq 0.04$ ”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。

例 7.1.2. 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.

在这个问题中, 设经过一段时间后罐装饮料容量 X 的平均值为 μ , 则由题意可设 $X \sim N(\mu, 10^2)$. 记 x_1, \dots, x_9 为取自这个正态总体 X 的一组样本观测值, 则 $\bar{x} = \frac{1}{9} \sum_{i=1}^9 x_i = 492$. 我们需要在“饮料平均容量为500 毫升”与“饮料平均容量为490 毫升”之间作判断, 即在“ $\mu = 500$ ”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**接受 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, \dots, 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**.

	事实	H_0 成立	H_1 成立
决策			
	接受 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)**，而“实际上 H_0 不成立但是它被接受”这样一类错误为**第II类错误(存伪)**。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，选定一个小的常数 α ，取 τ 使得犯第I类错误的概率，即 T 小于 τ 的概率小于 α 。称 α 为**显著性水平**。理想情况下， τ 取得恰好满足 $P_{H_0}(T < \tau) = \alpha$ 。为控制犯第I类错误的发生，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。设有假设检验问题

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1. \quad (7.1.1)$$

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**。构造一个适当的**检验统计量** $T = T(X_1, \dots, X_n)$ ，其中 X_1, \dots, X_n 为从总体中抽得的一个样本。根据对立假设的形状构造一个检验的**拒绝域** $W = \{T(X_1, \dots, X_n) \in A\}$ ，其中 A 为一个集合，通常是一个区间。比如拒绝域可取为 $\{T(X_1, \dots, X_n) > \tau\}$ ，则称 τ 为**临界值**。如果零假设成立但拒绝了零假设，则称犯了**第I类错误**，如果对立假设成立但接受零假设，则称犯了**第II类错误**。如对任意的 $\theta \in \Theta_0$ ，犯第I类错误的概率 $P_\theta(T(X_1, \dots, X_n) \in A)$ 小于或等于某个正的常数 α ，则称 α 为**显著性水平**。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。一个检

验 对应于一个拒绝域, 称 $\beta(\theta) = P_{\theta}(H_0 \text{ 被拒绝})$ 为检验的**功效函数**. 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.

§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. 如我国按照以前的司法制度, 公安机关抓到嫌疑犯后, 很多情况下要犯人自己证明无罪(有罪推断), 这对嫌疑犯很不利, 从而容易导致冤案. 现在的司法制度则总假定嫌疑犯是无罪的, 要司法部门证明其有罪(无罪推断), 这样做大大地有利于保护公民的利益, 如果要将真正的嫌疑犯绳之以法, 则司法部门必须有充分的证据, 这样做可以有效保护公民的权益, 对司法部门要求也变高了. 又比如药厂生产出一种新药, 在上市前要通过食品与药品监管局的检验. 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. 当然, 对立假设就是“新药比安慰剂效果好”. 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.

上述两原则是统一的: 一般不应该让受保护对象去证明一个命题.

§7.1.3 检验统计量的选取及假设检验的步骤

通过解答例7.1.1来说明假设检验的步骤.

例 7.1.3. (例 7.1.1续) 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?

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.

下面我们给出检验上述假设的一般步骤, 它的基本思想是: 一个好的点估计应该是一个优良检验的主要依据, 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 $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$, 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

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

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.