目录

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“$p \le 4\%$”，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。

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 $x_1, \cdots, x_9$ 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 $X_1, \cdots, X_n$, 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**.

事实 / 决策	H_0 成立	H_1 成立
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

$$H_0:\theta\in\Theta_0 \leftrightarrow H_1:\theta\in\Theta_1. \tag{7.1.1}$$

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** $T=T(X_1,\cdots,X_n)$, 其中 $X_1,\cdots,X_n$ 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** $W=\{T(X_1,\cdots,X_n)\in A\}$, 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**. 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 $P_\theta(T(X_1,\cdots,X_n)\in A)$ 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 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, $\beta(\theta)\le\alpha$. 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 $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$, 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.

7.2 重要参数检验

本节介绍最基本的假设检验问题: 一样本和两样本正态总体的有关均值和方差的检验, 简单的大样本检验 (0-1 分布参数的假设检验).

7.2.1 一样本正态总体均值和方差的检验

现实中经常碰到诸如此类的问题: 假设用于某用途的合格铁钉要求长度为 10 厘米, 现有经销商从生产厂家订购了一批这样的铁钉, 为了检验该批检验产品是否合格, 可以从中抽取一小部分进行测量检验, 通常铁钉的长度服从一个正态分布, 这类问题属于一样本正态总体的假设检验问题.

一般地, 设总体 $X \sim N(\mu, \sigma^2), -\infty < \mu < \infty, \sigma^2 > 0$; $X_1, \cdots, X_n$ 是取自总体 X 的一个样本. 取显著性水平为 α.

(1) 方差已知时均值的检验

先考虑双侧假设, 即要检验

$$H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0.$$

由于 μ 的极大似然估计为 $\bar{X}$, 取"标准化"后的检验统计量

$$U = u(X_1, \cdots, X_n) = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}$$

注意到当 H_0 成立时, $U \sim N(0,1)$, $|U|$ 应该较小, 反之当 $|U|$ 的观测值 $u(x_1, \cdots, x_n)$ 较大时, 不利于零假设 H_0 应该拒绝之. 所以选拒绝域形如

$$\{|U| > \tau\}.$$

要求显著性水平为 α, 即

$$P_{H_0}(|U| > \tau) = \alpha,$$

解得 $\tau = u_{\alpha/2}$. 于是检验的拒绝域为

$$\{|U| > u_{\alpha/2}\}.$$

即当观测值 $(x_1, \cdots, x_n)$ 满足不等式

$$\sqrt{n}\frac{|\bar{x} - \mu_0|}{\sigma} > u_{\alpha/2}$$

时拒绝 H_0.

类似地, 检验单侧假设

$$H_0:\ \mu=\mu_0 \leftrightarrow H_1:\ \mu>\mu_0 \quad \text{或者} \quad H_0:\ \mu\le\mu_0 \leftrightarrow H_1:\ \mu>\mu_0$$

仍然用统计量 U, 由于U 大时不利于 H_0, 取拒绝域为

$$\{U>u_\alpha\}.$$

而检验另一个单侧假设

$$H_0:\ \mu=\mu_0 \leftrightarrow H_1:\ \mu<\mu_0 \quad \text{或者} \quad H_0:\ \mu\le\mu_0 \leftrightarrow H_1:\ \mu<\mu_0$$

的拒绝域为

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

虽然我们取的临界值只考虑使检验在 $\mu=\mu_0$ 处的犯 I 类错误的概率为 α, 从检验的拒绝域的形状上可直接看出来在零假设下 $\mu\le\mu_0$ (或 $\mu\ge\mu_0$) 时犯第 I 类错误的概率恒小于或等于 α.

以上三个检验统称为 u 检验.

例 7.2.1. *随机地从一批铁钉中抽取 16 枚, 测得它们的长度 (单位: 厘米) 如下:*

2.942371 2.988662 3.106234 3.109316 3.118427 3.132254 3.140042 3.170188

2.902562 3.128003 3.146441 2.978240 3.103600 3.003394 3.044384 2.849916

已知铁钉长度服从标准差为 0.1 的正态分布, 在显著性水平 $\alpha=0.01$ 下, 能否认为这批铁钉的平均长度为 3 厘米? 如显著性水平为 $\alpha=0.05$ 呢?

解: 这是方差已知时关于均值 μ 的假设检验问题,

$$H_0:\mu=3 \leftrightarrow H_1:\mu\ne 3.$$

取检验统计量为 $U=\sqrt{n}(\bar{X}-3)/0.1$, 检验的拒绝域为 $|U|>u_{\alpha/2}$. 由样本算得检验统计量的值为 $u\approx 2.16$, 如显著性水平为 0.01, 则临界值为 $u_{0.005}\approx 2.58$, 跟检验统计量的值比较发现不能拒绝零假设, 即不能推翻铁钉平均长度为 3 厘米的假设; 而如果显著性水平为 0.05时, 临界值为 $u_{0.025}=1.96$, 此时可以拒绝零假设, 认为铁钉平均长度不等于 3 厘米. 这个例子说明结论可能跟显著性水平的选择有关: 显著性水平越小, 零假设被保护得越好从而更不容易被拒绝.

例 7.2.2. *对正态总体$N(\mu,\sigma^2)$(其中σ^2已知)下的假设检验问题 $H_0:\mu=\mu_0\leftrightarrow H_1:\mu\ne\mu_0$，如果我们还要求“犯第二类错误的概率要小于指定的$\beta>0$”该怎么办?*

解：根据功效函数和两类错误的定义，知道等价的要求

$$\beta_\phi(\mu) \geq 1-\beta, \quad \mu < \mu_0 \tag{7.2.1}$$

但是，当$\mu < \mu_0$但μ接近μ_0时，$\beta_\phi(\mu) \approx \alpha$，而因为$\alpha, \beta$一般都很小，因此一般有$\alpha < 1-\beta$，这就看出要求(7.2.1)无法达到。我们只能放松一些，要求对某个指定的$\mu_1 < \mu_0$，有

$$\beta_\phi(\mu) \geq 1-\beta, \quad \mu < \mu_1 \tag{7.2.2}$$

因为$\beta_\phi(\mu)$为μ的减函数，因此等价于要求

$$\beta_\phi(\mu_1) \geq 1-\beta$$

此即

$$\Phi\left(\frac{\sqrt{n}(\mu_0-\mu)}{\sigma} - u_\alpha\right) \geq 1-\beta$$

等价的得到

$$n \geq \sigma^2(u_\alpha + u_\beta)^2/(\mu_0-\mu)^2$$

也即要满足题目中的要求，样本大小至少要达到上式右边那么大。□

(2) 方差未知时均值的检验

考虑检验

$$H_0: \mu = \mu_0 \leftrightarrow \mu \neq \mu_0,$$

由于方差未知, 可以在将 $\bar{X}$ 标准化的过程中用样本方差 S^2 代替总体方差 σ^2, 得检验统计量

$$T = \sqrt{n}\frac{\bar{X}-\mu_0}{S}.$$

由于在 H_0 下, $T \sim t_{n-1}$, 于是拒绝域取成

$$\{|T| > t_{n-1}(\alpha/2)\}.$$

此检验称为 t 检验.

类似地可以得到另外两个单侧假设的检验拒绝域, 列于表 7.2.1中.

例 7.2.3. *(例7.2.1续) 设方差未知, 则在水平 0.01 和 0.05 下能否认为铁钉平均长度为 3 厘米?*

解: 这是方差未知时关于均值 μ 的假设检验问题,

$$H_0 : \mu = 3 \leftrightarrow H_1 : \mu \neq 3$$

取检验统计量为 $T = \sqrt{n}(\bar{X} - 3)/S$, 检验的拒绝域为 $|T| > t_{n-1}(\alpha/2)$. 由样本算得检验统计量的值约为 2.21, 与显著性水平 0.01 对应临界值 $t_{15}(0.005) \approx 2.95$ 比较, 不能拒绝零假设, 而与显著性水平 0.05 对应临界值 $t_{15}(0.025) \approx 2.13$ 比较, 可以拒绝零假设, 即在显著性水平 0.01 下不能拒绝铁钉平均长度为 3 厘米的假定, 但在显著性水平 0.05 下可以认为铁钉平均长度不等于 3 厘米, 此结论与方差已知情形一致.

(3) 方差的检验

考虑假设检验问题

$$H_0 : \sigma^2 = \sigma_0^2 \leftrightarrow H_1 : \sigma^2 \neq \sigma_0^2.$$

对均值已知的情形, 由 σ^2 的极大似然估计

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

可以构造检验统计量

$$\chi^2 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{n\hat{\sigma}^2}{\sigma_0^2}.$$

在 H_0 下, $\chi^2 \sim \chi_n^2$, χ^2 的平均值为 n, 而在 H_1 下, $\chi^2 = \frac{\sigma^2}{\sigma_0^2}\frac{n\hat{\sigma}^2}{\sigma^2}$ 的均值为 $\frac{\sigma^2}{\sigma_0^2}n \neq n$, 因此当 χ^2 的值过于偏离 n 时应该拒绝 H_0, 于是拒绝域取成

$$\left\{\chi^2 < \chi_n^2(1-\alpha/2) \text{ 或者 } \chi^2 > \chi_n^2(\alpha/2)\right\}.$$

对均值未知的情形, 构造检验统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2},$$

其中 S^2 为样本方差. 在 H_0 下, $\chi^2 \sim \chi_{n-1}^2$, 拒绝域取成

$$\left\{\chi^2 < \chi_{n-1}^2(1-\alpha/2) \text{ 或者 } \chi^2 > \chi_{n-1}^2(\alpha/2)\right\}.$$

对于单侧假设, 可以类似得到检验的拒绝域, 参看表 7.2.1.

上述检验称为 χ^2 检验.

例 7.2.4. *(例7.2.1续) 在水平 0.1 下能否认为铁钉的标准差大于 0.1 厘米?*

解: 这是均值未知时关于方差 σ^2 的假设检验问题,

$$H_0 : \sigma^2 \leq 0.1^2 \leftrightarrow H_1 : \sigma^2 > 0.1^2.$$

取检验统计量为 $\chi^2 = \frac{(n-1)S^2}{0.1^2}$, 检验的拒绝域为 $\{\chi^2 > \chi^2_{n-1}(\alpha)\}$. 由样本算得检验统计量的值 $\chi^2 \approx 14.32$, 与显著性水平 0.2 对应临界值 $\chi^2_{15}(0.1) \approx 22.31$ 比较, 不能拒绝零假设, 即在显著性水平 0.1 下可以认为铁钉的标准差小于 0.1.

表 7.2.1: 一样本正态总体 $N(\mu, \sigma^2)$.

检验对象	检验统计量	分布	拒绝域†
μ (σ^2已知)	$Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma$	$N(0,1)$	$\begin{cases} \lvert Z\rvert > u_{\alpha/2} \\ Z > u_\alpha \\ Z < -u_\alpha \end{cases}$
μ (σ^2 未知)	$T = \sqrt{n}(\bar{X} - \mu_0)/S$	t_{n-1}	$\begin{cases} \lvert T\rvert > t_{n-1}(\alpha/2) \\ T > t_{n-1}(\alpha) \\ T < -t_{n-1}(\alpha) \end{cases}$
σ^2 (μ 已知)	$\chi^2 = \frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i - \mu)^2$	χ^2_n	$\begin{cases} \chi^2 > \chi^2_n(\alpha/2)\text{或者}\chi^2 < \chi^2_n(1-\alpha/2) \\ \chi^2 > \chi^2_n(\alpha) \\ \chi^2 < \chi^2_n(1-\alpha) \end{cases}$
σ^2 (μ 未知)	$\chi^2 = \frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i - \bar{X})^2$	χ^2_{n-1}	$\begin{cases} \chi^2 > \chi^2_{n-1}(\alpha/2)\text{或者}\chi^2 < \chi^2_{n-1}(1-\alpha/2) \\ \chi^2 > \chi^2_{n-1}(\alpha) \\ \chi^2 < \chi^2_{n-1}(1-\alpha) \end{cases}$

†有关均值的检验: 对立假设分别为 $\mu \neq \mu_0$, $\mu > \mu_0$ 和 $\mu < \mu_0$. 有关方差的检验: 对立假设分别为 $\sigma^2 \neq \sigma_0^2$, $\sigma^2 > \sigma_0^2$ 和 $\sigma^2 < \sigma_0^2$.

表 7.2.1 总结了有关一样本正态总体的假设检验.

7.2.2 两样本正态总体的情形

为了检验某肥料是否能显著提高玉米产量, 可以设计一个随机试验: 选择两块条件一样的试验区, 把两试验区各分成若干小块, 一个试验区的各小块施肥, 另一个试验区的各小块不施肥, 最后统计收成, 可以采用如下的检验方法来检验玉米产量差别, 从而知道肥料是否有效.

设总体 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $-\infty < \mu_1, \mu_2 < \infty, \sigma_1^2, \sigma_2^2 > 0$; $X_1, \cdots, X_n$ 是从总体 X 中抽取的一个样本, $Y_1, \cdots, Y_n$ 是从总体 Y 中抽取的一个样本. 设来自不同总体的样本相互独立. 下面设考虑有关均值差 $\mu_1 - \mu_2$ 和方差比 σ_1^2/σ_2^2 的检验. 取显著性水平为 α. 举例说明.

例 7.2.5. *甲乙两个农业试验区种植玉米, 除了甲区施磷肥外, 其他试验条件都相同, 把两个试验区分别均分成 10 个和 9 个小区统计产量 (单位: 千克) , 得数据如下*

甲区 62 57 65 60 63 58 57 60 60 58

乙区 50 59 56 57 58 57 56 55 57

假定甲乙两区中每小块的玉米产量分别服从 $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$, 其中 μ_1, μ_2, σ^2 未知. 试问在显著性水平 $\alpha = 0.1$ 下磷肥对玉米的产量是否有效?

解: 磷肥对玉米产量有效果等价于 $\mu_1 > \mu_2$, 故将其设为对立假设, 假设检验问题是

$$H_0 : \mu_1 \leq \mu_2 = 0 \leftrightarrow H_1 : \mu_1 > \mu_2.$$

构造基于 $\mu_1 - \mu_2$ 的极大似然估计 $\bar{X} - \bar{Y}$ 的检验统计量

$$T = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{\frac{1}{m} + \frac{1}{n}}}.$$

当 H_0 成立时, $T \sim t_{m+n-2}$, 于是拒绝域为

$$\{T > t_{m+n-2}(\alpha)\}.$$

由所得数据算得检验统计量 T 的观测值为

$$t = \frac{\bar{x} - \bar{y}}{s_w \sqrt{\frac{1}{m} + \frac{1}{n}}} = 3.23.$$

由 $\alpha = 0.1$ 得临界值为 $t_{m+n-2}(\alpha/2) = t_{17}(0.1) \approx 1.33 < 3.23$, 因此拒绝 H_0, 即可以在显著性水平 0.1 下认为磷肥对玉米的产量有显著性影响.

例 7.2.6. *在例7.2.5中假定了两个正态总体的方差是相等的, 即 $\sigma_1^2 = \sigma_2^2 = \sigma^2$. 现在我们根据样本来检验这个方差齐性的假设, 即要检验*

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \leftrightarrow H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1.$$

解: 因为 σ_1^2 和 σ_2^2 的极大似然估计分别是

$$\hat{\sigma}_1^2 = \frac{1}{m} \sum_{i=1}^{m} (X_i - \bar{X})^2, \hat{\sigma}_2^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

在 $\theta=\sigma_1^2/\sigma_2^2$ 的极大似然估计 $\hat\theta=\hat\sigma_1^2/\hat\sigma_2^2$ 的基础上可以构造检验统计量

$$F=\frac{S_1^2}{S_2^2}=\frac{(m-1)\hat\sigma_1^2/m}{(n-1)\hat\sigma_2^2/n}.$$

注意到 F 中的分子和分母分别是 X 和 Y 的样本方差. 当零假设成立时, $F\sim F_{m-1,n-1}$. 于是拒绝域为

$$\{F<F_{m-1,n-1}(\alpha/2)\quad 或\quad F>F_{m-1,n-1}(1-\alpha/2)\}.$$

由数据算得检验统计量 F 的观测值 $f=1.19$, 如果取显著性水平 $\alpha=0.2$, 那么临界值为 $F_{9,8}(0.1)=2.44$, $F_{9,8}(0.9)=1/F_{8,9}(0.1)=0.41$ (如果 $X\sim F_{m,n}$, 则 $1/X\sim F_{n,m}$). 易见 $0.41<1.19<2.44$, 因此不能拒绝 H_0, 即在显著性水平 0.2 下可以认为上例中所作的方差齐性假定是合理的.

表 7.2.2 总结了两样本正态总体的双侧假设检验.

表 7.2.2: 两样本正态总体的假设检验

检验对象	检验统计量	分布	拒绝域†
均值(方差已知)	$Z=\frac{\bar X-\bar Y}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$	$N(0,1)$	$\begin{cases}\|Z\|>u(\alpha/2)\\ Z>u(\alpha)\\ Z<-u(\alpha)\end{cases}$
均值(方差未知)‡	$T=\frac{\bar X-\bar Y}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$	t_{m+n-2}	$\begin{cases}\|T\|>t_{m+n-2}(\alpha/2)\\ T>t_{m+n-2}(\alpha)\\ T<-t_{m+n-2}(\alpha)\end{cases}$
方差(均值已知)	$F=\frac{\sum_{i=1}^m(X_i-\mu_1)^2/m}{\sum_{i=1}^n(X_i-\mu_2)^2/n}$	$F_{m,n}$	$\begin{cases}F>F_{m,n}(\alpha/2)或F<\frac{1}{F_{n,m}(\alpha/2)}\\ F>F_{m,n}(\alpha)\\ F<\frac{1}{F_{n,m}(\alpha)}\end{cases}$
方差(均值未知)	$F=\frac{S_1^2}{S_2^2}$	$F_{m-1,n-1}$	$\begin{cases}F>F_{m-1,n-1}(\alpha/2)或F<\frac{1}{F_{n-1,m-1}(\alpha/2)}\\ F>F_{m-1,n-1}(\alpha)\\ F<\frac{1}{F_{n-1,m-1}(\alpha)}\end{cases}$

†有关均值的检验: 对立假设分别为 $\mu_1\neq\mu_2$, $\mu_1>\mu_2$ 和 $\mu_1<\mu_2$. 有关方差的检验: 对立假设分别为 $\sigma_1^2\neq\sigma_2^2$, $\sigma_1^2>\sigma_2^2$ 和 $\sigma_1^2<\sigma_2^2$.
‡假定方差相等

7.2.3 成对数据

在上述两样本正态总体的假设检验中, 要求两个样本是独立的, 但是没有要求样本量相等. 有一类数据叫做成对数据 $\{(X_1,Y_1),\cdots,(X_n,Y_n)\}$, 比如一个病人在用药前后测得的指标分别为 X 和 Y, 则 X 与 Y 总是一起出现的, 且由于它们是同一个体的指标, 故具有很大的相关性而绝对不是独立的, 这与两样本正态总体有本质区别. 另外, 两样本检验问题要求样本 $X_1,\cdots,X_m$ 是同分

布的 $(Y_1,\cdots,Y_n$ 亦然), 而成对数据则无此要求, 而要求 $X_1-Y_1,\cdots,X_n-Y_n$ 是同分布. 比如病人可以是来自两个不同性别、种族、年龄层的人. 要检验用药前后的指标有无显著差别, 可以构造一个新的总体 $Z=Y-X$ 及样本 $Z_1=X_1-Y_1,\cdots,Z_n=X_n-Y_n$, 相应的假设检验是一样本的! 在实际问题中, 如果发现有两个样本且其样本量是相等的, 则要检查独立性和同分布性, 否则可能是成对数据.

7.2.4 0-1 分布中未知参数 p 的假设检验

产品验收时, 需要检验不合格率是否小于某给定的一个数.

设 $(X_1,\cdots,X_n)$ 是取自总体 X 的一个样本, 该总体服从 0-1 分布, 取 1 的概率为 p. 常见的假设有三种:

(1) $H_0: p=p_0 \leftrightarrow H_1: p\neq p_0$;

(2) $H_0: p=p_0 \leftrightarrow H_1: p>p_0$ 或 $H_0: p\leq p_0 \leftrightarrow H_1: p>p_0$;

(3) $H_0: p=p_0 \leftrightarrow H_1: p<p_0$ 或 $H_0: p\geq p_0 \leftrightarrow H_1: p<p_0$.

假定样本量 n 较大, 取显著性水平为 α. 由于 p 的极大似然估计为 $\bar{X}$, 取"标准化"过的检验统计量

$$T=\sqrt{n}\frac{\bar{X}-p_0}{\sqrt{p_0(1-p_0)}},$$

其中 p_0 和 $p_0(1-p_0)/n$ 分别为 $\bar{X}$ 在零假设 $p=p_0$ 下的期望和方差, 从而当 H_0 成立时, 由中心极限定理近似地有 $T\sim N(0,1)$. 于是上述三种检验的拒绝域分别为

$$\{|T|>u_{\alpha/2}, \quad \{T>u_\alpha\} \quad \text{和} \quad \{T<-u_\alpha\}$$

例 7.2.7. *某厂产品不合格率通常为 0.5. 厂方希望知道原料产地的改变是否对产品的质量发生显著的影响. 现在随机地从原料产地改变后的产品中抽取了 80 个样品进行检验, 发现有 5 个是不合格品. 试问, 在显著性水平 0.1 下, 厂方由此可以得出什么结论?*

解: 总体 $X\sim B(1,p)$, 其中 p 未知. 在显著性水平 $\alpha=0.1$ 下, 产品质量无变化等价于 $p=0.05$, 故我们要检验

$$H_0: p=0.05 \leftrightarrow H_1: p\neq 0.05.$$

由于 $\bar{x}=5/80=0.0625$, 因此检验统计量 T 的观测值

$$t=\sqrt{n}\frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)}}=0.513.$$

由 $\alpha = 0.10$ 得临界值 $u_{0.05} = 1.645$. 易见, $|t| < 1.645$, 因此不能拒绝 H_0, 即在近似显著性水平 0.10 下可以认为原料产地的改变对该厂产品的质量没有发生显著的影响.

7.2.5 置信区间和假设检验之间的关系

置信区间和假设检验之间有着明显的联系。我们首先考虑置信区间和双边检验之间的关系。设$X_1, \cdots, X_n$为从总体$F(x;\theta)$ 中抽取的样本，参数θ的$1-\alpha$置信区间为$[\underline{\theta}, \bar{\theta}]$,即

$$P(\underline{\theta} \leq \theta \leq \bar{\theta}) \geqq 1-\alpha$$

而对假设$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$，在原假设之下，有

$$P(\underline{\theta} \leq \theta_0 \leq \bar{\theta}) \geqq 1-\alpha$$

等价于

$$P(\theta_0 > \bar{\theta}) + P(\theta_0 < \underline{\theta}) \leq \alpha$$

按显著性检验的定义，即得其检验为

$$\phi : \text{当 } \underline{\theta} \leq \theta_0 \leq \bar{\theta} \text{ 时，接受} H_0, \text{不然就拒绝}$$

反过来讲,如果假设$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$检验的接受域有形式

$$\underline{\theta}(x_1, \cdots, x_n) \leq \theta_0 \leq \bar{\theta}(x_1, \cdots, x_n)$$

即有

$$P(\underline{\theta} \leq \theta_0 \leq \bar{\theta}) \geqq 1-\alpha$$

由θ_0的任意性，知对任意的θ，有

$$P(\underline{\theta} \leq \theta \leq \bar{\theta}) \geqq 1-\alpha$$

即：为求出参数θ的$1-\alpha$置信区间，我们可以先找出θ的双边检验 $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$ 的检验函数，则其接受域就是参数θ的$1-\alpha$置信区间。反过来，为求假设 $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$的检验，我们可以先求出参数$\theta$的$1-\alpha$置信区间，则就是该假设的接受域。

类似地，置信水平为$1-\alpha$的单侧置信区间$(\underline{\theta}, \infty)$ (或者$(-\infty, \bar{\theta})$)与显著性水平为α的右(或者左)边检验问题 $H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$ (或者$H_0 : \theta \geq \theta_0 \leftrightarrow H_1 : \theta < \theta_0$)，也有类似的对应关系。

7.3 拟合优度检验

前面的假设检验基本上是在假定总体是正态的条件下做的, 但是这个假设本身不一定成立, 需要收集样本 $(X_1, \cdots, X_n)$ 来检验它. 一般地, 检验

$$H_0 : X\text{服从某种分布}$$

可以采用 Karl Pearson 提出的 χ^2 拟合优度检验.

7.3.1 离散总体情形

(1) 理论分布不含未知参数的情形

设某总体 X 服从一个离散分布, 且根据经验得知总体落在类别 $a_1, \cdots, a_k$ 的理论频率分别为 $p_1, \cdots, p_k$, 现从该总体抽得一个样本量为 n 的样本, 其落在类别 $a_1, \cdots, a_k$ 的观测数分别为 $n_1, \cdots, n_k$. 感兴趣的问题是检验理论频率是否正确, 即下面假设是否正确:

$$H_0 : P(X \in a_1) = p_1, \cdots, P(X \in a_k) = p_k.$$

这类问题只提零假设而不提对立假设, 相应的检验方法称为拟合优度检验. 显然, 在零假设下, 各类别的理论频数分别为 $np_1, \cdots, np_k$, 将理论频数和观测频数列于下表:

类别	a_1	a_2	$\cdots$	a_k
理论频数	np_1	np_2	$\cdots$	np_k
观测频数	n_1	n_2	$\cdots$	n_k

由大数定律知, 在零假设成立时, n_i/n 依概率收敛于 p_i, 故理论频数 np_i 与观测频数 n_i 接近. 而检验统计量取为

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}.$$

简单地, 就是

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

其中 O 为观测频数, E 为期望频数.

这个统计量中每项的分母的选取有点讲究, 我们可以这样粗略地解释: 假设 n_i 服从 Poisson 分布, 则 n_i 的均值和方差均为 np_i, 从而 $(n_i - np_i)/\sqrt{np_i}$ 的极限分布为标准正态分布, 因此 χ^2 近似为 k 个服从自由度为 1 的 χ^2 分布的随机变量之和, 由于 $\sum_{i=1}^{k}(n_i - np_i) = 0$, 故这 k 个随机变

量满足一个约束, 从而 χ^2 的自由度为 $k-1$. 事实上, 可以严格地证明, 在一定的条件下, χ^2 的极限分布就是自由度为 $k-1$ 的 χ^2 分布, 但其证明超出本课程的要求范围.

下面给出一个例子来说明拟合优度检验的应用.

例 7.3.1. *有人制造一个含 6 个面的骰子, 并声称是均匀的. 现设计一个实验来检验此命题: 连续投掷 600 次, 发现出现六面的频数分别为 97, 104, 82, 110, 93, 114. 问能否在显著性水平 0.2 下认为骰子是均匀的?*

解: 该问题设计的总体是一个有 6 个类别的离散总体, 记出现六个面的概率分别为 $p_1, \cdots, p_6$, 则零假设可以表示为

$$H_0 : p_i = 1/6, i = 1, \cdots, 6.$$

在零假设下, 理论频数都是 100, 故检验统计量 χ^2 的取值为

$$\frac{(97-100)^2}{100} + \frac{(104-100)^2}{100} + \frac{(82-100)^2}{100} + \frac{(110-100)^2}{100} + \frac{(93-100)^2}{100} + \frac{(114-100)^2}{100} = 6.94,$$

跟自由度为 $6-1=5$ 的 χ^2 分布的上 0.05 分位数 $\chi^2_5(0.2) \approx 7.29$ 比较, 不能拒绝零假设, 即可在显著性水平 0.2 下认为骰子是均匀的.

例 7.3.2. *孟德尔(Mendel)豌豆杂交试验。纯黄和纯绿品种杂交，因为黄色对绿色是显性的，在Mendel第一定律(自由分离定律)的假设下，二代豌豆中应该有75%是黄色的，25%是绿色的。在产生的$n=8023$个二代豌豆中，有$n_1 = 6022$个黄色，$n_2 = 2001$个绿色。我们的问题是检验这些这批数据是否支持Mendel第一定律，要检验的假设是*

$$H_0 : \quad \pi_1 = 0.75, \ \pi_2 = 0.25$$

解: 在Mendel第一定律(H_0)下，黄色和绿色的个数期望值为

$$\mu_1 = n\pi_1 = 8023 * 0.75 = 6017.25, \ \mu_2 = n\pi_2 = 8023 * 0.25 = 2005.75$$

则Pearson χ^2统计量为

$$Z = \sum \frac{(O-E)^2}{E} = (6022 - 6017.25)^2/6017.25 + (2001 - 2005.75)^2/2005.75 = 0.015$$

自由度$df = 1$，$p\text{−}value$为0.903. 因此可以认为这些数据服从Mendel第一定律。Fisher基于Mendel试验的所有数据，发现其数据与理论值符合的太好，$p-value = 0.99993$，但这么好的拟合在几万次试验中才发生一次，因而Fisher断定数据可能有伪造的嫌疑[注1]。

[注1]https://arxiv.org/pdf/1104.2975.pdf

(2) 理论分布含若干未知参数的情形

当理论总体总含有未知的参数时, 理论频数 np_i 一般也与这些参数有关, 此时应该用适当的估计如极大似然估计代替这些参数以得到 p_i 的估计 $\hat{p}_i$, 得到的统计量记为

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}.$$

拟合优度检验的提出者 Karl Pearson 最初认为在零假设下, 检验统计量的 χ^2 的极限分布仍等于自由度为 $k-1$ 的 χ^2 分布, R. A. Fisher 发现自由度应该等于 $k-1$ 减去估计的独立参数的个数 r, 即 $k-1-r$.

例 7.3.3. *从某人群中随机抽取 100 个人的血液, 并测定他们在某基因位点处的基因型. 假设该位点只有两个等位基因 A 和 a, 这 100 个基因型中 AA, Aa 和 aa 的个数分别为30, 40, 30, 则能否在 0.05 的水平下认为该群体在此位点处达到 Hardy-Weinberg 平衡态?*

解: 取零假设为

$$H_0: \text{Hardy-Weinberg 平衡态成立.}$$

设人群中等位基因 A 的频率为 p, 则该人群在此位点处达到 Hardy-Weinberg 平衡态指的是在人群中 3 个基因型的频率分别为 $P(AA) = p^2$, $P(Aa) = 2p(1-p)$ 和 $P(aa) = (1-p)^2$, 即零假设可等价地写成

$$H_0: P(AA) = p^2, P(Aa) = 2p(1-p), P(aa) = (1-p)^2.$$

在 H_0 下, 3 个基因型的理论频数为 $100 \times \hat{p}^2$, $100 \times 2 \times \hat{p}^2(1-\hat{p})$ 和 $100 \times (1-\hat{p})^2$, 其中 $\hat{p}$ 等于估计的等位基因频率 0.5, 代入 χ^2 统计量表达式, 得统计量的值等于 4. 该统计量的值大于自由度为 $3-1-1=1$ (恰好一个自由参数被估计) 的 χ^2 分布上 0.05 分位数 3.84, 故可在 0.05 的水平下认为未达到 Hardy-Weinberg 平衡态.

7.3.2 列联表的独立性和齐一性检验

(1) 独立性检验

下面考虑很常用的列联表. 列联表是一种按两个属性作双向分类的表. 例如肝癌病人可以按所在医院 (属性 A) 和是否最终死亡 (属性 B) 分类. 目的是看不同医院的疗效是否不同. 又如婴儿可按喂养方式 (属性 A, 分两个水平: 母乳喂养与人工喂养) 和小儿牙齿发育状况 (属性 B, 分两个水平: 正常与异常) 来分类. 这两个例子中两个属性都只有两个水平, 相应的列联表称为“四格表”, 一

般地, 如果第一个属性有 a 个水平, 第二个属性有 b 个水平, 称为 $a\times b$ 表 (见教材 p268) . 实际应用中, 常见的一个问题是考察两个属性是否独立. 即零假设是

$$H_0: \text{属性 A 与属性 B 独立.}$$

这是列联表的独立性检验问题.

假设样本量为 n, 第 (i,j) 格的频数为n_{ij}. 记 $p_{ij}=P(\text{属性 A, B 分别处于水平}i,j)$, $u_i=P(\text{属性 A 有水平}i)$, $v_i=P(\text{属性 B 有水平}j)$. 则零假设就是 $p_{ij}=u_iv_j$. 将 u_i 和 v_j 看成参数, 则总的独立参数有 $a-1+b-1=a+b-2$ 个. 它们的极大似然估计为

$$\hat{u}_i=\frac{n_{i\cdot}}{n},\hat{v}_j=\frac{n_{\cdot j}}{n}.$$

正好是它们的频率 (证明参看教材) . 其中 $n_{i\cdot}=\sum_{j=1}^b n_{ij}$, $n_{\cdot j}=\sum_{i=1}^a n_{ij}$. 在 H_0 下, 第 (i,j) 格的理论频数为 $n\hat{p}_{ij}=n_{i\cdot}n_{\cdot j}/n$, 因此在 H_0 下, $\sum_{i=1}^a\sum_{j=1}^b(n_{ij}-n\hat{p}_{ij})$ 应该较小. 故取检验统计量为

$$\chi^2 \quad = \quad \sum_{i=1}^a\sum_{j=1}^b\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n)^2}{(n_{i\cdot}n_{\cdot j}/n)}.$$

在零假设下 χ^2 的极限分布是有自由度为 $k-1-r=ab-1-(a+b-2)=(a-1)(b-1)$ 的 χ^2 分布. 对于四格表, 自由度为 1.

(2) 齐一性检验

跟列联表有关的另一类重要的检验是齐一性检验, 即检验某一个属性 A 的各个水平对应的另一个属性 B 的分布全部相同, 这种检验跟独立性检验有着本质的区别. 独立性问题中两属性都是随机的; 而齐一性问题中属性 A 是非随机的, 这样涉及到的分布实际上是条件分布. 虽然如此, 所采用的检验方法跟独立性检验完全一样.

例 7.3.4. *下面表是甲乙两医院肝癌病人生存情况. 需要根据这些数据判断两医院的治疗效果是否一样.*

甲、乙两院肝癌的近期疗效

	生存	死亡	合计
甲院	$150(n_{11})$	$88(n_{12})$	$238(n_{1\cdot})$
乙院	$36(n_{21})$	$18(n_{22})$	$54(n_{2\cdot})$
合计	$186(n_{\cdot 1})$	$106(n_{\cdot 2})$	$292(n)$

解: 这是一个齐一性检验问题. 检验统计量 χ^2 的观测值为 0.2524, 远远小于自由度为 1 的 χ^2 分布的上 0.05 分位数, 故可以接受零假设, 即在水平 0.05 下可以认为两个医院的疗效无差别的.

当有某个格子的频数较小时, 如果允许的话可以合并格子是每个格子的频数足够大, 实际问题中不允许合并格子 (合并后失去了实际意义), 此时可以用 **Fisher 的精确检验法**.

7.3.3 连续总体情形

设 $(X_1, \cdots, X_n)$ 是取自总体 X 的一个样本, 记 X 的分布函数为 $F(x)$, 需要检验的那种分布中含有 r 个总体参数 $\theta_1, \cdots, \theta_r$. 我们要在显著性水平 α 下检验

$$H_0 : F(x) = F_0(x; \theta_1, \cdots, \theta_r),$$

其中 $F_0(x; \theta_1, \cdots, \theta_r)$ 表示需要检验的那种分布的分布函数. 例如, 当我们要检验

$$H_0 : X \sim N(\mu, \sigma^2)$$

时, $r = 2$, $\theta_1 = \mu, \theta_2 = \sigma^2$.

$$F_0(x; \mu, \sigma^2) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\} dt.$$

上述假设可以通过适当的离散化总体分布, 采用拟合优度法来做检验. 首先把实数轴分成 k 个子区间 $(a_{j-1}, a_j]$, $j = 1, \cdots, k$, 其中 a_0 可以取 $-\infty$, a_k 可以取 ∞. 这样构造了一个离散总体, 其取值就是这 k 个区间. 记

$$p_j \quad = \quad P_{H_0}(a_{j-1} < X \leq a_j) = F_0(a_j; \theta_1, \cdots, \theta_r) - F_0(a_{j-1}; \theta_1, \cdots, \theta_r), j = 1, \cdots, k.$$

如果 H_0 成立, 则概率 p_j 应该与数据落在区间 $(a_{j-1}, a_j]$ 的频率 $f_j = n_j/n$ 接近, 其中 n_j 表示相应的频数. 当 p_i 的取值不含未知参数时, 取检验统计量

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j},$$

否则取

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - n\hat{p}_j)^2}{n\hat{p}_j},$$

其中 $\hat{p}_i$ 是将 p_i 中的未知参数换成适当的估计后得到的 p_i 的估计. 拒绝域取为

$$\left\{\chi^2 > \chi^2_{k-r-1}(\alpha)\right\}.$$

如果 p_i 中不含未知参数, 则 $r = 0$.

使用 χ^2 进行拟合优度检验时一般要求 $n \geq 50$, $n\hat{p}_j \geq 5, j = 1, \cdots, k$, 如果不满足这个条件, 最好把某些组作适当合并.

例 7.3.5. *从某连续总体中抽取一个样本量为 100 的样本, 发现样本均值和样本标准差分别为 -0.225 和 1.282, 落在不同区间的频数如下表所示:*

区间	$(-\infty, -1)$	$[-1, -0.5)$	$[-0.5, 0)$	$[0, 0.5)$	$[0.5, 1)$	$[1, \infty)$
观测频数	*25*	*10*	*18*	*24*	*10*	*13*
理论频数	*27*	*14*	*16*	*14*	*12*	*17*

可否在显著性水平 0.05 下认为该总体服从正态分布?

解: 设理论正态分布的均值和方差分别为 μ 和 σ^2, 记第 i 个区间为 $(a_{i-1}, a_i,\ i = 1, \cdots, 6$, 则样本落在第 i 个格子的理论概数为 $100P(a_{i-1} < X \leq a_i)$, 其中 $X \sim N(\mu, \sigma^2)$. 将 $\mu = -0.225$ 和 $\sigma^2 = \frac{99}{100} \times 1.282^2 = 1.622$ 代入得到估计的理论频数, 列于上表中.

$$H_0: \text{总体服从正态分布}$$

由此算得检验统计量 χ^2 的值约为 9.25, 与自由度为 6-1-2=3 的 χ^2 分布的上 0.05 分位数 $\chi_3^2(0.05) \approx 7.81$ 比较可以拒绝零假设, 即可以在显著性水平 0.05 下认为该总体不服从正态分布.