Lec9: 假设检验

张伟平

2011 年 4 月 11 日

假设检验即使用样本对所关心的假设进行推断. 假设检验问题大致分为两大类:

1. 参数型假设检验: 即总体的分布形式已知(如正态、指数、二项分布等) , 总体分布依赖于未知参数(或参数向量) θ, 要检验的是有关未知参数的假设. 如$X \sim N(a, \sigma^2)$, a未知, 检验

$$H_0: \ a = a_0 \longleftrightarrow H_1: a \neq a_0 \ \text{或} \ H_0: \ a \leq a_0 \longleftrightarrow H_1: a > a_0$$

2. 非参数型假设检验: 如果总体分布形式未知, 此时就需要有一种与总体分布族的具体数学形式无关的统计方法, 称为非参数方法. 如检验一批数据是否来自某个已知的总体, 就属于这类问题.

1 假设检验的若干基本概念

一、检验问题的提法

为了说明假设检验问题的提法, 考察下面的例子.

例5.1.1 某工厂生产的一大批产品, 要卖给商店. 按规定次品率p不得超过0.01,今在其中抽取100件, 经检验有3件次品, 问这批产品可否出厂?

关于这个问题, 在我们面前存在两种可能性:

$$\text{甲: } 0 < p \leq 0.01; \quad \text{乙: } 0.01 < p < 1.$$

我们要通过从这批产品中抽样来决定甲, 乙两种可能性中哪个成立.

这个问题常以下述方式提出: 引进一个“假设”

$$H_0: \ 0 < p \leq 0.01$$

它叫做零假设 (null hypothesis) 或原假设, 有时也简称为假设. 另一个可能是

$$H_1: \ 0.01 < p < 1$$

叫做对立假设或备择假设(alternative hypothesis).

我们的目的是要通过样本决定接受H_0, 还是拒绝H_0. 可以形象地把问题写成

$$H_0: \ 0 < p \leq 0.01 \longleftrightarrow H_1: \ 0.01 < p < 1$$

注意这个提法中将H_0放在中心位置, 它是检验的对象. H_0和H_1的位置不可颠倒. 从这个例子可将假设检验问题一般化, 提法如下:

设有参数分布族$\{F_\theta, \theta \in \Theta\}$,此处$\Theta$为参数空间. $X_1, \cdots, X_n$是从上述分布族中抽取的简单随机样本. 在参数假设检验问题中, 我们感兴趣的是θ是否属于参数空间Θ的某个真子集Θ_0, 则命题$H_0: \ \theta \in \Theta$称为零假设或原假设, 其确切含义是: 存在一个$\theta_0 \in \Theta_0$使得X的分布为F_{θ_0}. 记$\Theta_1 = \Theta - \Theta_0$,则命题$H_1: \ \theta \in \Theta_1$称为$H_0$的对立假设或备择假设(在例5.1.1中,$\Theta = (0,1)$, $\Theta_0 = (0, 0.01]$, $\Theta_1 = (0.01, 1)$) . 则假设检验问题表为

$$H_0: \ \theta \in \Theta_0 \longleftrightarrow H_1: \ \theta \in \Theta_1, \tag{1.1}$$

在(5.1.1)式中, 若Θ_0或Θ_1只包含参数空间Θ中的一个点, 则称为简单假设 (simple hypothesis);否则, 称为复合假设 (composite hypothesis). 例如, 样本抽自$N(a, \sigma_0^2)$, σ_0^2已知, 则参数空间为$\Theta = \{a : -\infty < a < +\infty\}$. 令$H_0 : a = a_0 \longleftrightarrow H_1 : a \neq a_0$,则$H_0$为简单假设, H_1为复合假设. 再如, 同上问题, 令$H_0' : a \leq a_0 \longleftrightarrow H_1' : a > a_0$, 则零假设$H_0'$和对立假设$H_1'$皆为复合假设.

二、假设检验的依据–小概率原理

在例5.1.1中, 由于这批产品数量很大, 故若记X为抽取的100件产品中的次品数, 则可以近似认为$X \sim B(100, p)$. 如果零假设$0 < p \leq 0.01$是正确的, 则

$$P(X \geq 3) \leq 1 - \sum_{i=0}^{2} C_{100}^{i} 0.01^{i} 0.99^{100-i} = 0.079$$

即如果认为这批产品是合格的, 则100件产品中有3件次品或者更多次品的可能性只有7.9%, 这个概率比较小, 按照小概率原理, 不大可能在一次实验中就发生, 但我们偏偏观测到了. 因此有理由怀疑零假设是不正确的.

应用小概率原理只能大体上表达我们对零假设是否成立的大致推断.

三、否定域、检验函数和检验统计量

我们仍通过例子来说明这个概念.

例5.1.2 设$\mathbf{X} = (X_1, \cdots, X_n)$为从总体$X \sim N(a, 1)$中抽取的随机样本. 考虑检验问题:

$$H_0: \ a = a_0 \longleftrightarrow H_1: \ a \neq a_0, \tag{1.2}$$

此处, a_0为给定的常数.

这种检验的一种直观上的作法是: 先求a的一个估计量,我们知道$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$是$a$的一个优良估计. 若$|\bar{X} - a_0|$较大,我们就倾向于否定$H_0$;反之, 如果$|\bar{X} - a_0|$较小, 我们就认为抽样结

果与H_0相接近,因而倾向于接受H_0. 具体地说,我们要确定一个数A,对样本$X=(X_1,\dots,X_n)$,算出$\bar{X}$, 当$|\bar{X}-a_0|>A$时就否定H_0;当$|\bar{X}-a_0|\le A$时就接受H_0.我们称

$$D=\{\mathbf{X}=(X_1,\dots,X_n):\ |\bar{X}-a_0|>A\} \tag{1.3}$$

为否定域,或叫做拒绝域 (reject region).即,否定域是由样本空间 $\mathcal{X}$中一切使$|\bar{X}-a_0|>A$的那些样本$\mathbf{X}=(X_1,\dots,X_n)$构成. 有了否定域,等价于将样本空间 $\mathscr{X}$ 分成不相交的两部分$\mathcal{X}_1=\mathcal{X}-D$和$\mathcal{X}_2=D$,一旦有了样本$\mathbf{X}$,当$\mathbf{X}\in\mathcal{X}_1$时,就接受$H_0$; 当$\mathbf{X}\in\mathcal{X}_2=D$时,就否定$H_0$. 我们称$\mathcal{X}_1$为接受域 (acceptance region).只要A定下来了,则否定域(或接受域) 也就确定了. 因此, 在此问题中检验可视为如下的一种法则:

$$T:\begin{cases}\text{当 } |\bar{X}-a_0|>A \text{ 时}, & \text{拒绝 } H_0\\ \text{当 } |\bar{X}-a_0|\le A \text{ 时}, & \text{接受 } H_0\end{cases}$$

上式中的T,给定了一种法则,一旦有了样本,我们就可以在接受H_0或否定H_0 这两个结论中选择一个. 我们称这样一种法则T为检验问题(1.2)的一个检验.

为了便于数学上处理,我们引入如下检验函数$\varphi(\mathbf{x})$的概念,$\varphi(\mathbf{x})$与检验T 是一一对应的.在例5.1.2中

$$\varphi(\mathbf{x})=\begin{cases}1 & \text{当 } |\bar{X}-a_0|>A \text{ 时}\\ 0 & \text{当 } |\bar{X}-a_0|\le A \text{ 时}\end{cases} \tag{1.4}$$

我们有如下定义:

定义5.1.1 由(5.1.4)给出的检验函数 $\varphi(\mathbf{x})$是定义在样本空间 $\mathscr{X}$上, 取值于[0,1]的函数. 它表示当有了样本X后,否定H_0的概率.

由定义可见,若$\varphi(\mathbf{x})=1$,则以概率为1否定H_0,当$\varphi(\mathbf{x})=0$, 则以概率为0否定H_0(即以概率为1接受H_0).若$\varphi(\mathbf{x})$只取0,1这两个值, 则这种检验称为非随机化检验 (non-randomized test).此时,否定域也可用检验函数表示如下: $D=\{\mathbf{X}=(X_1,\dots,X_n):\ \varphi(\mathbf{x})=1\}$.

若对某些样本$\mathbf{X}$,有$0<\varphi(\mathbf{x})<1$, 则称$\varphi(\mathbf{x})$为随机化检验 (randomized test).如在例5.1.1中,设$\mathbf{X}=(X_1,\dots,X_n)$为样本,当$\sum_{i=1}^{100}X_i<c$ 时认为这批产品合格,接受H_0;当$\sum_{i=1}^{100}X_i>c$时,认为不合格, 拒绝H_0.当$\sum_{i=1}^{100}X_i=c$时,若规定拒绝H_0,厂方觉得被拒绝的可能性大了, 吃亏了. 反之, 若接受H_0,买方(商店)接受不合格产品的可能性大了,也觉得吃亏.在双方僵持不下的情况下,下列折中方案是双方都可以接受的: 定下一个数$0<r<1$,规定当$\sum_{i=1}^{100}X_i=c$时,以概率为r作一次试验,根据试验结果来决定拒绝还是接受这批产品. 如取$r=1/2$,则可通过掷一枚硬币来决定.规定若出现正面则拒绝H_0, 否则接受H_0.这样,当出现$\sum_{i=1}^{100}X_i=c$,双方都有1/2的可能,做出对自己不利的决定,双方都觉得合理, 可以接受.如果取$r=1/3$,试验可以通过在有2个白球和1个黑球的盒子中摸球来决定, 若摸到黑球(发生的概率为1/3)则拒绝H_0,若摸到白球(发生的概率为2/3) 则接受H_0.这种随机化检验函数可表为

$$\varphi(\mathbf{x})=\begin{cases}1 & \text{若 } \sum_{i=1}^{100}X_i>c\\ r & \text{若 } \sum_{i=1}^{100}X_i=c\\ 0 & \text{若 } \sum_{i=1}^{100}X_i<c.\end{cases} \tag{1.5}$$

在例5.1.2中要确定检验,必须定出(5.1.3)或(5.1.4)式中的A, 此处A 称为临界值 (critical value).要定下c的值需要找到检验统计量的分布. 在此例中检验统计量是$T = \bar{X}$. 同样在例5.1.1中, 检验函数(5.1.5) 中的c称为临界值,检验统计量是$T = \sum_{i=1}^{100} X_i$.确定检验统计量的分布是解决假设检验问题的关键. 当检验统计量的精确分布很难找到时, 若其极限分布比较简单,我们可用极限分布代替精确分布,获得假设检验问题的近似解.

四、两类错误与功效函数

统计推断是以样本为依据的,由于样本的随机性,我们不能保证统计推断方法的绝对正确性,而只能以一定的概率去保证这种推断的可靠性.在假设检验问题中可能出现下列两种情形会犯错误:

决策 \ 假设	拒绝H_0	接受H_0
H_0为真	犯错	不犯错
H_1为真	不犯错	犯错

1. 零假设 H_0本来是对的, 由于样本的随机性, 样本观察值落入否定域D, 错误地将H_0否定了, 称为弃真. 这时犯的错误称为第一类错误 (Type I error).

2. 零假设 H_0本来不对, 由于样本的随机性, 样本观察值落入接受域$\bar{D}$, 错误地将H_0接受了, 称为取伪. 这时犯的错误称为第二类错误(Type II error).

如在例5.1.1中确定了非随机检验如下:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & 若 \sum_{i=1}^{100} X_i > 3; \\ 0 & 若 \sum_{i=1}^{100} X_i \le 3. \end{cases}$$

如果总体的真实次品率为$p = 0.005 < 0.01$,由于样本的随机性,抽样结果显示$\sum_{i=1}^{100} X_i = 5$,即样本落入否定域,这时我们犯第一类错误. 但也有可能总体的真实次品率$p = 0.03 > 0.01$, 由于样本的随机性,抽样结果显示$\sum_{i=1}^{100} X_i = 1$,即样本落入了接受域. 这时我们犯第二类错误.

应当注意,在每一具体场合,我们只会犯两类错误中的一个. 当检验确定后,犯两类错误的概率也就确定了. 我们希望犯两类错误的概率越小越好,但这一点很难做到. 在样本大小n固定的前提下,二者不可兼得. 这就如同区间估计问题中可靠度和精度二者不可兼得一样. 那么,怎样去计算犯两类错误的概率呢?为此, 引出功效函数的概念.

定义5.1.2 设$\varphi(\mathbf{x})$是$H_0: \theta \in \Theta_0 \longleftrightarrow H_1: \theta \in \Theta_1$的一个检验函数,则

$$\beta_\varphi(\theta) = P_\theta\big\{用检验\ \varphi\ 否定了\ H_0\big\} = E_\theta[\varphi(\mathbf{X})], \quad \theta \in \Theta$$

称为φ的功效函数 (power function),也称为效函数或势函数.

若$\varphi(\mathbf{x})$为非随机化检验,否定域为D,则

$$\beta_\varphi(\theta) = P_\theta\big(\mathbf{X} = (X_1, \ldots, X_n) \in D\big)$$

因此功效函数表示当样本分布参数为θ时,否定H_0的概率. 对例5.1.1,当检验函数为随机化检验(1.5)时, 利用$\sum_{i=1}^n X_i \sim b(n,\theta)$, $0<\theta<1$可知检的功效函数为

$$\begin{aligned}\beta_\varphi(\theta) = E_\theta[\varphi(\mathbf{x})] &= P\left(\sum_{i=1}^n X_i > c\right) + rP\left(\sum_{i=1}^{100} X_i = c\right) \\ &= \sum_{k=c+1}^{100} \binom{100}{k} \theta^k (1-\theta)^{100-k} + r\binom{100}{c}\theta^c(1-\theta)^{100-c}.\end{aligned}$$

以下讨论中假定$\varphi(\mathbf{x})$皆为非随机化的检验函数,除非特别申明, 不认为$\varphi(\mathbf{x})$为随机化检验函数.

知道了检验$\varphi(\mathbf{x})$的功效函数后,就可以计算犯两类错误的概率. 若以$\alpha^*_\varphi(\theta)$ 和$\beta^*_\varphi(\theta)$分别记犯第一、二类错误的概率,则犯第一类错误的概率可表示为

$$\alpha^*_\varphi(\theta) = \begin{cases} \beta_\varphi(\theta) & \text{当 } \theta \in \Theta_0 \\ 0 & \text{当 } \theta \in \Theta_1, \end{cases}$$

犯第二类错误的概率可表示为

$$\beta^*_\varphi(\theta) = \begin{cases} 0 & \text{当 } \theta \in \Theta_0 \\ 1-\beta_\varphi(\theta) & \text{当 } \theta \in \Theta_1. \end{cases}$$

还需要说明的一点是: 如前所述, 犯两类错误的概率完全由功效函数决定, 从这一点上看, 如果两个检验有同一功效函数, 则此两检验在性质上也完全相同.

四、检验水平和控制犯第一类错误概率的原则

前面说过, 我们希望一个检验犯两类错误的概率都很小, 但除极例外情形, 一般说来在固定样本大小时对任何检验都办不到. 例如, 要使犯第一类错误的概率减小, 就要缩小拒绝域, 使接受域增大, 这必然导致犯第二类错误概率增大, 反之亦然. 因此, Neyman-Pearson提出了一条原则, 就是限制犯第一类错误概率的原则. 即在保证犯第一类错误的概率不超过指定数值α $(0<\alpha<1$,通常取较小的数) 的检验中, 寻找犯第二类错误概率尽可能小的检验. 若记

$$S_\alpha = \{\,\varphi:\ \alpha^*_\varphi(\theta) = \beta_\varphi(\theta) \le \alpha,\ \text{当}\theta \in \Theta_0\},$$

S_α表示由所有犯第一类错误的概率都不超过α的检验函数构成的类. 我们只考虑S_α中的检验. 在S_α中挑选“犯第二类错误的概率尽可能小的检验”, 这种法则称为控制犯第一类错误概率的法则.

> 根据Neyman-Pearson原则, 在原假设H_0为真时, 我们作出错误决定(即否定H_0) 的概率受到了控制. 这表明, 原假设H_0受到保护, 不致于轻易被否定. 所以在具体问题中, 我们往往将有把握、不能轻易否定的命题作为原假设H_0, 而把没有把握的、不能轻易肯定的命题作为对立假设. 因此原假设H_0和对立假设H_1的地位是不平等的, 不能相互调换.

与犯第一类错误概率相联系的另一个概念是检验水平, 其定义如下:

定义5.1.3 设φ是(1.1)的一个检验, 而$0 \le \alpha \le 1$. 如果φ犯第一类错误的概率总不超过α(或等价地说, 若φ满足: $\beta_\varphi(\theta) \le \alpha$,对一切$\theta \in \Theta_0$) , 则称$\alpha$是检验$\varphi$的一个水平, 而$\varphi$称为显著性水平为$\alpha$的检验, 简称水平为$\alpha$的检验.

按这一定义,检验的水平不唯一.若α为检验φ的水平, 而$\alpha < \alpha' < 1$,则α'也是检验φ的水平. 为避免这一问题, 有时称一个检验的最小水平为其真实水平. 也就是

$$\text{检验}\varphi\text{的真实水平} = sup\{\ \beta_\varphi(\theta),\ \theta \in \Theta_0\} \tag{1.6}$$

至于水平的选择,习惯上把α取得比较小且标准化,如$\alpha = 0.01, 0.05, 0.10$等. 标准化是为了方便造表.

水平的选取, 对检验的性质有很大影响. 不难了解,如果水平选得很低, 那么我们容许犯第一类错误的概率很小,而为了达到这一点势必大大缩小否定域, 而这样就增加了犯第二类错误的可能性. 反之,若水平选得高,则否定域扩大, 使接受域缩小, 从而犯第二类错误的概率相应的将降低. 这样看来,水平的选择不是一个数学问题,而是一个必须从实际角度来考虑的问题. 一般说来有以下几个因素影响水平的选定.

1. 当一个检验涉及两方利益时, 水平的选定常是双方协议的结果. 以例5.1.1为例,商店向工厂进货, 检验其次品率是否超过0.01,若水平选的低,则可能有较多的次品被商店接受;反之,若水平定的高, 则将有较多的合格品被商店拒收. 因此水平定的大小涉及商店和工厂双方利益,应由双方商定. 如前所述,有时还要采取随机化的方法,使双方利益达到平衡.

2. 两种错误的后果一般在性质上有很大的不同. 如果第一类错误的后果在性质上很严重, 我们就力求在合理的范围内尽量减少犯这种错误的可能性,这时相应的水平就取得更低一些. 例如,制药厂要生产一种新药代替旧药治疗某种疾病, 安排了一些试验,要对新旧药物疗效作出检验. 由于旧药已经长期临床使用,有一定的疗效. 新药尚未经长期临床使用, 一旦效果不好时,将危及病人的生命安全, 造成的后果会很严重. 所以在进行检验时, 将原假设H_0设为"旧药不比新药差",且使检验水平α定得更小一些, 这样使H_0被否定的可能性大大减小了. 这样就保证了: "原假设被否定、新药被接受的检验" 将是非常严格的.

3. 一般说来,试验者在试验前对问题的情况总不是一无所知的. 他对问题的了解使他对零假设 是否能成立就有了一定的看法, 这种看法可能影响到他对水平的选择. 比方说一个物理学家根据某种理论推定随机变量X应有分布F,而他打算将这一理论付诸检验. 很明显,如果他对这一理论很有信心,他将非常倾向于认为假设能成立, 这时只有很有力的证据才可能使他认为这假设不对. 相应地,他将把检验水平取得低一些.

在实际问题中, 零假设被否定, 常常意味着推翻一种理论或用新方法来代替一直使用的标准方法. 在大多数情况下,人们希望这样做时有相当大的根据. 从这里可以看到, Neyman-Pearson控制犯第一类错误的原则,在零假设的选择中有很大的实际意义, 而决不单纯是一数学问题. 同时,也进一步理解了在假设检验问题中,零假设处在突出地位的原因.

最后要说明的一点是:若水平α很小,原假设H_0不会轻易被否定. 如果样本观察值落入了否定域, 我们做出"否定原假设H_0"的结论就比较可靠(因为, 此时我们只会犯第一类错误, 且其概率很小). 反之, 当α很小时,如果样本观察值落入接受域,我们做出"接受原假设H_0"的结论未必可靠. 这只能表明:在所选定的水平下没有充分根据认为H_0不成立,决不意味着有充分根据说明它正确(因为此时我们只会犯第二类错误, 但其概率可能很大) .

五、求解假设检验问题的一般步骤

1. 根据问题的要求提出零假设H_0和备择假设H_1;

2. 导出否定域的形式, 确定检验统计$T(\mathbf{X})$, 其中临界值A待定.

3. 选取适当水平, 利用检验统计量的分布求出临界值A.

4. 由样本$\mathbf{X}$算出检验统计量$T(\mathbf{X})$的具体值, 代入到否定域中, 与临界值相比较, 作出接受或者拒绝原假设H_0的结论.

2 正态总体参数的假设检验

正态分布是最常见的分布, 关于它的参数的假设检验是实际中常遇到的问题, 因此也是最重要的一类检验问题. 本节将分下列几种情况来讨论正态总体参数的直观检验方法: 单个正态总体均值和方差的检验; 两个正态总体均值差和方差比的检验; 极限分布为正态分布的有关大样本检验.

在讨论正态分布总体参数的假设检验问题时, §2.4中的定理2.2.3和推论2.4.2–推论2.4.5在求检验统计量的分布中起到十分重要的作用.

一、单个正态总体均值的检验

设$\mathbf{X}=(X_1,\ldots,X_n)$为从正态总体$N(\mu,\sigma^2)$中抽取的简单随机样本, 给定检验水平$\alpha$, 求下列三类检验问题:

(1) $H_0:\ \mu=\mu_0 \longleftrightarrow H_1:\ \mu\neq\mu_0$;

(2) $H_0':\ \mu\le\mu_0 \longleftrightarrow H_1':\ \mu>\mu_0$;

(3) $H_0'':\ \mu\ge\mu_0 \longleftrightarrow H_1'':\ \mu<\mu_0$;

其中μ_0和检验水平α给定.

我们称检验问题(1)为*双边检验* (two-side test), 称检验问题(2)和(3)为*单边检验* (one-side test).

单个正态总体方差未知时的检验问题要比方差已知情况更常见, 我们将重点讨论这一情形. 关于均值已知时的检验方法, 我们将在后面给一个说明.

首先考虑检验问题(1), 即

$$H_0:\ \mu=\mu_0 \longleftrightarrow H_1:\ \mu\neq\mu_0$$

我们用直观方法构造检验的否定域. 我们知道$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 是μ的无偏估计, 且具有良好性质. 直观上看$|\bar{X}-\mu_0|$越大, H_0越不象成立. 因此检验的否定域可取如下形式: $\{\mathbf{X}=(X_1,\ldots,X_n):\ |\bar{X}-\mu_0|>A\}$, A待定. 当σ^2未知时, 由推论2.4.2可知,在$\mu=\mu_0$条件下,

$$T=\frac{\sqrt{n}(\bar{X}-\mu_0)}{S}\sim t_{n-1}. \tag{2.1}$$

因此取$T=\frac{\sqrt{n}(\bar{X}-\mu_0)}{S}$ 作为检验统计量,则否定域的等价形式可取为

$$\big\{(X_1,\ldots,X_n):\ |T|>c\big\},\ c\ 待定.$$

由检验水平为α,可知

$$P(|T|>c\,|\,H_0)=P\left(\left|\sqrt{n}(\bar{X}-\mu_0)/S\right|>c\,\middle|\,H_0\right)=\alpha,$$

故得$c=t_{n-1}(\alpha/2)$.因此由否定域

$$D_1=\left\{(X_1,\ldots,X_n):\ \left|\sqrt{n}(\bar{X}-\mu_0)/S\right|>t_{n-1}(\alpha/2)\right\} \tag{2.2}$$

确定的检验为检验问题(1)的水平为α的检验.

对检验问题(2), 取$T=\sqrt{n}(\bar{X}-\mu_0)/S$作为检验统计量,因此直观上否定域的形式为

$$\{(X_1,\ldots,X_n):\ T>c\},\ c\text{ 待定}.$$

为使检验$(2')$具有水平α,, 即要求

$$P(T>c\,|\,H_0')=P\left(\sqrt{n}(\bar{X}-\mu_0)/S>c\,\middle|\,\mu\le\mu_0\right)\le\alpha, \tag{2.3}$$

注意到

$$\begin{aligned}P\left(\sqrt{n}(\bar{X}-\mu_0)/S>c\,\middle|\,\mu\le\mu_0\right) &= P\left(\sqrt{n}(\bar{X}-\mu)/S>c+\frac{\sqrt{n}(\mu_0-\mu)}{S}\,\middle|\,\mu\le\mu_0\right)\\ &\le P\left(\sqrt{n}(\bar{X}-\mu)/S>c,\,\middle|\,\mu\le\mu_0\right)\end{aligned}$$

因此只需$P\left(\sqrt{n}(\bar{X}-\mu)/S>c,\,\middle|\,\mu\le\mu_0\right)=\alpha$则式(2.3)成立. 所以$c=t_{n-1}(\alpha)$.故检验问题$(2')$的水平为$\alpha$的检验的否定域是

$$D_2^*=\left\{(X_1,\ldots,X_n):\ T=\sqrt{n}(\bar{X}-\mu_0)/S>t_{n-1}(\alpha)\right\}. \tag{2.4}$$

类似方法可得检验问题(3)的水平为α的检验的否定域.

$$D_3=\left\{(X_1,\ldots,X_n):\sqrt{n}(\bar{X}-\mu_0)/S<-t_{n-1}(\alpha)\right\}$$

这种基于检验统计量服从t分布的检验方法称为一样本t检验 ,T称为一样本t检验统计量.

注5.2.1 方差已知时正态总体均值的检验方法作如下说明: 在$N(\mu,\sigma^2)$中, 当σ^2已知, 且H_0成立, 即$\mu=\mu_0$时有

$$U=\sqrt{n}(\bar{X}-\mu_0)/\sigma\sim N(0,1) \tag{2.5}$$

因此取U=$\sqrt{n}(\bar{X}-\mu_0)/\sigma$ 作为检验统计量, 用完全与方差σ^2未知情形相同的方法(所不同的就是用检验统计量U 代替那儿的检验统计量T) 可得检验问题(1)-(3)的水平为α的检验的否定域. 否定域中的临界值将t_{n-1}分位数改成相应的标准正态分布的分位数. 详细结果见表5.2.1.

这种基于检验统计量服从$N(0,1)$分布的检验方法称为一样本U检验.

表5.2.1 单个正态总体均值的假设检验

	H_0	H_1	检验统计量及其分布	否定域
σ^2	$\mu=\mu_0$	$\mu\ne\mu_0$	$U=\sqrt{n}(\bar{X}-\mu_0)/\sigma$	$\lvert U\rvert>u_{\alpha/2}$
已	$\mu\le\mu_0$	$\mu>\mu_0$	$U\lvert\mu=\mu_0\sim N(0,1)$	$U>u_\alpha$
知	$\mu\ge\mu_0$	$\mu<\mu_0$		$U<-u_\alpha$
σ^2	$\mu=\mu_0$	$\mu\ne\mu_0$	$T=\sqrt{n}(\bar{X}-\mu_0)/s$	$\lvert T\rvert>t_{n-1}(\alpha/2)$
未	$\mu\le\mu_0$	$\mu>\mu_0$	$T\lvert\mu=\mu_0\sim t_{n-1}$	$T>t_{n-1}(\alpha)$
知	$\mu\ge\mu_0$	$\mu<\mu_0$		$T<-t_{n-1}(\alpha)$

例5.2.1 食品厂用自动装罐机装罐头食品,每罐标准重量为500克,每天开工需检查机器的工作状况. 今抽得10罐,测得其重量(单位:克)

$$495,\ 510,\ 505,\ 498,\ 503,\ 492,\ 502,\ 512,\ 497,\ 506.$$

假定罐头重量X服从正态分布$N(\mu,\sigma^2)$,已知$\sigma = 6.5$, 问机器是否工作正常?(取检验水平$\alpha = 0.05$)

解 检验问题为

$$H_0:\ \mu = 500 \longleftrightarrow H_1:\ \mu \neq 500$$

本题中σ^2已知,故否定域由表5.2.1中第一行给出,即

$$D = \left\{(X_1,\ldots,X_n):\ \left|\tfrac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma}\right| > \mu_{\frac{\alpha}{2}}\right\}$$

其中$n = 10,\ \alpha = 0.05$,查表得$u_{0.025} = 1.96$,由样本算得$\bar{X} = 502$, 因此

$$|U| = \left|\tfrac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma}\right| = \left|\tfrac{\sqrt{10}(502-500)}{6.5}\right| = 0.973 < 1.96$$

所以在显著性水平$\alpha = 0.05$下认为没有足够理由说明自动装罐机工作不正常,故接受H_0.

例5.2.2 某砖厂所生产的地砖的抗断强度X服从正态分布$N(\mu,\sigma^2)$, 今从该厂生产的地砖中随机抽取6块测得抗断强度如下(单位:kg/cm^2)

$$32.56,\ 29.66,\ 31.64,\ 30.00,\ 31.87,\ 31.03.$$

问这一批地砖的抗断强度可否认为不低于32.50 kg/cm^2? (取检验水平$\alpha = 0.05$)

解 检验问题为

$$H_0:\ \mu \geq 32.50 \longleftrightarrow H_1:\ \mu < 32.50$$

本题中σ^2未知,故采用t检验法,否定域由表5.2.1中最后一行给出,即

$$D = \{(X_1,\ldots,X_n):\ \tfrac{\sqrt{n}(\bar{X}-\mu_0)}{S} < -t_{n-1}(\alpha)\},$$

其中$n = 6,\ \alpha = 0.05$,查表得$t_5(0.05) = 2.015$,由数据算得$\bar{X} = 31.13,\ S = 1.123$,因此有

$$T = \tfrac{\sqrt{6}(31.13-32.50)}{1.123} = -3.36 < -2.015$$

故否定H_0,即认为地砖强度达不到32.50 kg/cm^2.

二、单个正态总体方差的检验

设$X_1,\ldots,X_n$为自正态总体$N(\mu,\sigma^2)$中抽取的简单随机样本, 本段要讨论下列三类检验问题:

(4) $H_0:\ \sigma^2 = \sigma_0^2 \longleftrightarrow H_1:\ \sigma^2 \neq \sigma_0^2;$

(5) $H_0':\ \sigma^2 \leq \sigma_0^2 \longleftrightarrow H_1':\ \sigma^2 > \sigma_0^2;$

(6) $H_0'':\ \sigma^2 \geq \sigma_0^2 \longleftrightarrow H_1'':\ \sigma^2 < \sigma_0^2;$

其中σ_0^2和检验水平α给定.

我们将重点讨论, 均值未知时单个正态总体时方差的检验方法. 当均值已知时如何处理, 我们将在后面给一个说明. 首先考虑检验问题(4), 即

$$H_0:\ \sigma^2 = \sigma_0^2 \longleftrightarrow H_1:\ \sigma^2 \neq \sigma_0^2$$

由于均值μ未知, 我们知道$S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$ 是σ^2的一个无偏估计, 且具有良好性质. 直观上看S^2/σ_0^2太小或者S^2/σ_0^2太大时,H_0不象成立. 因此检验的否定域可取如下形式: $\{(X_1,\ldots,X_n): S^2/\sigma_0^2<A_1$ 或 $S^2/\sigma_0^2>A_2\}$, A_1, A_2 待定. 在给定$\sigma^2=\sigma_0^2$的条件下, 由定理2.2.3可知

$$(n-1)S^2/\sigma_0^2\sim\chi_{n-1}^2. \tag{2.6}$$

故取检验统计量为$\chi^2=(n-1)S^2/\sigma_0^2$. 因此, 否定域的等价形式可取为

$$D=\left\{(X_1,\ldots,X_n):\ (n-1)S^2/\sigma_0^2<c_1 \text{ 或 } (n-1)S^2/\sigma_0^2>c_2\right\}$$

记$\theta=(\mu,\sigma^2)$,为了确定c_1, c_2,令

$$\alpha=P_\theta\left((n-1)S^2/\sigma_0^2<c_1 \text{ 或 } (n-1)S^2/\sigma_0^2>c_2\middle|H_0\right)$$

满足上式的c_1, c_2的对子有很多,存在一对c_1, c_2是最优的,但计算较复杂, 且使用不方便. 确定c_1, c_2的一个简单实用的方法是:令

$$P_\theta\left((n-1)S^2/\sigma_0^2<c_1\middle|H_0\right)=\alpha/2,\qquad P_\theta\left((n-1)S^2/\sigma_0^2>c_2\middle|H_0\right)=\alpha/2$$

由上述两式和(2.6)易知临界值$c_1=\chi_{n-1}^2(1-\alpha/2)$, $c_2=\chi_{n-1}^2(\alpha/2)$.所以检验问题(4)的水平为$\alpha$的接受域为

$$\bar{D}_4=\left\{(X_1,\ldots,X_n):\ \chi_{n-1}^2(1-\alpha/2)\le(n-1)S^2/\sigma_0^2\le\chi_{n-1}^2(\alpha/2)\right\}$$

此接受域的表达式比否定域简单,使用上也方便, 故此处采用接受域代替否定域.

用完全类似于(一) 中求检验问题(2)、(3)的方法可分别求得检验问题(5)和(6)的水平为α的否定域如下:

$$D_5=\left\{(X_1,\ldots,X_n):\ (n-1)S^2/\sigma_0^2>\chi_{n-1}^2(\alpha)\right\},$$
$$D_6=\left\{(X_1,\ldots,X_n):\ (n-1)S^2/\sigma_0^2<\chi_{n-1}^2(1-\alpha)\right\}.$$

注5.2.2 当均值μ已知时, 方差σ^2的检验方法简述如下: 当μ已知时,σ^2的一个具有良好性质的无偏估计是$S_*^2=\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^2$.当$\sigma^2=\sigma_0^2$时,由推论2.4.1 可知

$$nS_*^2/\sigma_0^2=\sum_{i=1}^n(X_i-\mu)^2/\sigma_0^2\sim\chi_n^2. \tag{2.7}$$

因此取检验统计量为$\chi_*^2=nS_*^2/\sigma_0^2$,用它代替$\chi^2=(n-1)S^2/\sigma_0^2$.采用完全类似于前面$\mu$未知情形的讨论方法, 可得检验问题(4)-(6)的水平为α的否定域,仅注意在否定域中将确定临界值的χ^2分布的自由度由$n-1$改成n即可. 详细结果见表5.2.2.

这种基于检验统计量服从一定自由度的χ^2分布的检验方法称为χ^2检验.

表5.2.2 单个正态总体方差的假设检验

	H_0	H_1	检验统计量及其分布	否定域
μ 已知	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi_*^2=ns_*^2/\sigma_0^2$	$nS_*^2/\sigma_0^2<\chi_n^2(1-\alpha/2)$ 或 $nS_*^2/\sigma_0^2>\chi_n^2(\alpha/2)$
	$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$	$\chi_*^2\vert\sigma_0^2\sim\chi_n^2$	$nS_*^2/\sigma_0^2>\chi_n^2(\alpha)$
	$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$		$nS_*^2/\sigma_0^2<\chi_n^2(1-\alpha)$
μ 未知	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\frac{(n-1)s^2}{\sigma_0^2}$	$(n-1)S^2/\sigma_0^2<\chi_{n-1}^2(1-\alpha/2)$ 或 $(n-1)S^2/\sigma_0^2>\chi_{n-1}^2(\alpha/2)$
	$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$	$\chi^2\vert\sigma_0^2\sim\chi_{n-1}^2$	$(n-1)S^2/\sigma_0^2>\chi_{n-1}^2(\alpha)$
	$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$		$(n-1)S^2/\sigma_0^2<\chi_{n-1}^2(1-\alpha)$

例5.2.3 某工厂生产的一种细纱支数服从正态分布,其总体标准差为1.2. 现从某日生产的一批产品中抽取16缕进行支数测量,测得样本标准差为2.1,问纱的均匀度是否变劣?($\alpha=0.05$)

解 检验问题为

$$H_0:\ \sigma^2=1.44\longleftrightarrow H_1:\ \sigma^2\neq1.44.$$

检验的接受域为

$$D=\{(X_1,\ldots,X_n):\chi_{n-1}^2(1-\alpha/2)<(n-1)S^2/\sigma_0^2<\chi_{n-1}^2(\alpha/2)\}$$

此处$n=16$, $\alpha=0.05$.查表得到$\chi_{15}^2(0.975)=6.262$, $\chi_{15}^2(0.025)=27.488$. 由已知数据算得$S^2=2.1^2=4.41$, $\sigma_0^2=1.2^2=1.44$.因此有

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}=\frac{15\times4.41}{1.44}=45.94>27.488.$$

故否定H_0,即认为棉纱的均匀度变劣.

三、两个正态总体均值差的检验

设$\mathbf{X}=(X_1,\ldots,X_m)$为自正态总体$N(\mu_1,\sigma_1^2)$ 中抽取的简单随机样本, $\mathbf{Y}=(Y_1,\ldots,Y_n)$为自正态总体$N(\mu_2,\sigma_2^2)$ 中抽取的简单随机样本, 且合样本$X_1,\ldots,X_m;Y_1,\ldots,Y_n$独立. 本段要讨论下列三类检验问题:

(7) $H_0:\ \mu_2-\mu_1=\mu_0\longleftrightarrow H_1:\ \mu_2-\mu_1\neq\mu_0$;

(8) $H_0':\ \mu_2-\mu_1\leq\mu_0\longleftrightarrow H_1':\ \mu_2-\mu_1>\mu_0$;

(9) $H_0'':\ \mu_2-\mu_1\geq\mu_0\longleftrightarrow H_1'':\ \mu_2-\mu_1<\mu_0$;

其中μ_0和检验水平α给定.

以上假设检验问题与求均值差的置信区间的*Behrens-Fisher*问题 (见§4.2,三) 一样, 除了它的几种特殊情形已获得圆满解决外, 一般情况至今尚没有得到简单、精确的解法. 下面分几种情况分别讨论.

1. 当σ_1^2和σ_2^2已知时, 均值差的检验问题

首先考虑单边检验问题(7), 即

$$H_0:\ \mu_2-\mu_1\leq\mu_0\longleftrightarrow H_1:\ \mu_2-\mu_1>\mu_0$$

由于$\bar{Y}-\bar{X}$是$\mu_2-\mu_1$的一个无偏估计, 且具有良好性质. 直观上看$|\bar{Y}-\bar{X}-\mu_0|$越大时, H_0越不象成立, 故否定域可取如下形式$\{(X_1,\dots,X_m;Y_1,\dots,Y_n):|\bar{Y}-\bar{X}-\mu_0|>A\}$, A待定. 由于$\bar{Y}-\bar{X}\sim N(\mu_2-\mu_1,\sqrt{\sigma_1^2/m+\sigma_2^2/n})$, 故当$\mu_2-\mu_1=\mu_0$时有

$$U=\frac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\ \sim\ N(0,1). \tag{2.8}$$

因此, 取检验统计量为$U=(\bar{Y}-\bar{X}-\mu_0)\big/\sqrt{\sigma_1^2/m+\sigma_2^2/n}$, 则否定域等价形式为

$$\left\{(X_1,\dots,X_m;Y_1,\dots,Y_n):\ \left|\frac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\right|>c\right\},\ \ c\text{ 待定.}$$

记$\theta=\mu_2-\mu_1$, 为确定c,令

$$\alpha=P_\theta\left(|U|>c|H_0\right)=P\left(\left|\frac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\right|>c\middle|H_0\right)=2-2\Phi(c)$$

由此可确定临界值$c=u_{\alpha/2}$.因此检验问题(3)的水平为α的检验否定域为

$$D_7=\left\{(X_1,\dots,X_m;Y_1,\dots,Y_n):\ \left|\frac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\right|>u_{\alpha/2}\right\},$$

完全类似(一)中的求检验问题(2)、(3)方法, 可得检验问题(8)、(9)的水平为α的检验的否定域为

$$D_8=\left\{(X_1,\dots,X_m;Y_1,\dots,Y_n):\ \frac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}>u_\alpha\right\}$$
$$D_9=\left\{(X_1,\dots,X_m;Y_1,\dots,Y_n):\ \frac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}<-u_\alpha\right\}$$

这种基于由(2.8)给出的检验统计量U的检验方法,称为两样本U检验.

2. 当$\sigma_1^2=\sigma_2^2=\sigma^2$未知时, 均值差的检验

若$\sigma_1^2=\sigma_2^2=\sigma^2$已知, 则由前面刚刚讨论的两样本$U$检验可知, 此时检验统计量$U$变为

$$U=\frac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma^2/m+\sigma^2/n}}=\frac{\bar{Y}-\bar{X}-\mu_0}{\sigma}\sqrt{\frac{mn}{m+n}}\,. \tag{2.9}$$

而在$\sigma_1^2=\sigma_2^2=\sigma^2$未知时, 上述表达式中的$\sigma^2$常用

$$\begin{aligned}S_w^2&=\frac{1}{n+m-2}\left((m-1)S_1^2+(n-1)S_2^2\right)\\&=\frac{1}{n+m-2}\left(\sum_{i=1}^m(X_i-\bar{X})^2+\sum_{j=1}^n(Y_j-\bar{Y})^2\right)\end{aligned} \tag{2.10}$$

来估计, 其中

$$S_1^2=\frac{1}{m-1}\sum_{i=1}^m(X_i-\bar{X})^2,\qquad S_2^2=\frac{1}{n-1}\sum_{j=1}^n(Y_j-\bar{Y})^2$$

分别为样本$X_1,\ldots,X_m$和$Y_1,\ldots,Y_n$的样本方差. 当$\mu_2-\mu_1=\mu_0$时,将公式(2.9)中的σ用(2.10)中的s_w代替, 得到下列的统计量T_w,由推论2.4.3可知

$$T_w=\frac{\bar{Y}-\bar{X}-\mu_0}{S_w}\sqrt{\frac{mn}{m+n}}\sim t_{n+m-2}. \tag{2.11}$$

因此我们取$T_w=\frac{\bar{Y}-\bar{X}-\mu_0}{S_w}\sqrt{\frac{mn}{m+n}}$作为检验统计量.

与(一)中一样本t检验方法类似, 在两样本均值差的检验问题中, 用检验统计量T_w 代替那儿的一样本t检验统计量T,用完全相同的讨论方式, 可得检验问题(7)-(9)的水平为α 的否定域, 只要注意在否定域中将确定临界值的t分布的自由度由$n-1$改为$n+m-2$即可. 详细结果见下列表5.2.3.

这种基于检验统计量服从t_{n+m-2}分布的检验方法,称为两样本t检验.

表5.2.3 两个正态总体均值差的假设检验

<table>
<tr><th></th><th>H_0</th><th>H_1</th><th>检验统计量
及其分布</th><th>否定域</th></tr>
<tr><td rowspan="3">σ_1^2
σ_2^2
已
知</td><td>$\mu_2-\mu_1=\mu_0$</td><td>$\mu_2-\mu_1\neq\mu_0$</td><td rowspan="3">$U=\frac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}$
$U|\mu_0\sim N(0,1)$</td><td>$|U|>u_{\frac{\alpha}{2}}$</td></tr>
<tr><td>$\mu_2-\mu_1\leq\mu_0$</td><td>$\mu_2-\mu_1>\mu_0$</td><td>$U>u_\alpha$</td></tr>
<tr><td>$\mu_2-\mu_1\geq\mu_0$</td><td>$\mu_2-\mu_1<\mu_0$</td><td>$U<-u_\alpha$</td></tr>
<tr><td rowspan="3">σ_1^2
σ_2^2
未
知</td><td>$\mu_2-\mu_1=\mu_0$</td><td>$\mu_2-\mu_1\neq\mu_0$</td><td rowspan="3">$T_w=\frac{\bar{Y}-\bar{X}-\mu_0}{S_w}\sqrt{\frac{mn}{m+n}}$
$T_w|\mu_0\sim t_{n+m-2}$
$S_w^2=\frac{(m-1)S_1^2+(n-1)S_2^2}{n+m-2}$</td><td>$|T_w|>$
$t_{n+m-2}(\frac{\alpha}{2})$</td></tr>
<tr><td>$\mu_2-\mu_1\leq\mu_0$</td><td>$\mu_2-\mu_1>\mu_0$</td><td>$T>t_{n+m-2}(\alpha)$</td></tr>
<tr><td>$\mu_2-\mu_1\geq\mu_0$</td><td>$\mu_2-\mu_1<\mu_0$</td><td>$T<-t_{n+m-2}(\alpha)$</td></tr>
</table>

例5.2.4 为研究正常成年男女血液红细胞平均数的差别, 检验某地正常成年男子156人, 女子74人, 计算男女红细胞的平均数和样本标准差分别为

$$\text{男: } \bar{X}=465.13\text{ 万 }/mm^3,\quad S_1=54.80\text{ 万 }/mm^3$$
$$\text{女: } \bar{Y}=422.16\text{ 万 }/mm^3,\quad S_2=49.20\text{ 万 }/mm^3$$

假定正常成年男女红细胞数分别服从正态分布, 且方差相同. 检验正常成年人红细胞数是否与性别有关. $(\alpha=0.01)$

解 设$X_1,\ldots,X_m$ i.i.d. $\sim N(\mu_1,\sigma^2)$; $Y_1,\ldots,Y_n$ i.i.d. $\sim N(\mu_2,\sigma^2)$, 且假定这两组样本独立. 检验问题为

$$H_0:\ \mu_2-\mu_1=\mu_0\longleftrightarrow H_1:\ \mu_2-\mu_1\neq\mu_0$$

检验的否定域为

$$D=\left\{(X_1,\ldots,X_m,Y_1,\ldots,Y_n):\left|\frac{\bar{Y}-\bar{X}}{S_w}\sqrt{\frac{mn}{m+n}}\right|>t_{n+m-2}(\tfrac{\alpha}{2})\right\}$$

此处

$$m=156,\ n=74,\ \bar{X}=465.13,\ \bar{Y}=422.16,\ S_1=54.80,\ S_2=49.20,$$
$$S_w^2=\frac{1}{n+m-2}\left[(m-1)S_1^2+(n-1)S_2^2\right]=2816.6,\quad S_w=53.07,$$

查表得$t_{228}(0.005)=2.576$.由

$$|T|=\left|\frac{\bar{Y}-\bar{X}}{S_w}\sqrt{\frac{mn}{m+n}}\right|=\left|\frac{422.16-465.13}{53.07}\sqrt{\frac{156\times 74}{156+74}}\right|=5.74>2.576,$$

故否定H_0,即认为正常成年人的红细胞数与性别有关.

3. 当样本容量$m=n$时均值差的检验—成对比较问题

前面讨论的用于两个正态总体均值差的检验中, 假定了来自两个正态总体的样本是相互独立的. 但在实际问题中, 有时候情况不总是这样. 可能这两个正态总体的样本是来自同一个总体上的重复观察, 它们是成对出现的, 而且是相关的. 例如为了观察一种安眠药的效果, 记录了n 个失眠病人服药前的每晚睡眠时间$X_1,X_2,\ldots,X_n$和服用此安眠药后每晚睡眠时间$Y_1,Y_2,\ldots,Y_n$.这里(X_i,Y_i)是第i个病人不服用安眠药和服用安眠药每晚的睡眠时间. 它们是有关系的, 不会相互独立. 另一方面, $X_1,X_2,\ldots,X_n$是n个不同失眠病人的睡眠时间, 由于个人体质诸方面的条件不同, 这n个观察值不能认为是来自同一个正态总体的样本. $Y_1,Y_2,\ldots,Y_n$也是一样. 这样的数据称为成对数据. 这样的数据用两样本t检验就不合适, 因为X_i和Y_i是同在第i个病人身上观察到的夜晚睡眠时间, 所以$Z_i=Y_i-X_i$ 就消除了人的体质诸方面的差异, 仅剩下安眠药的效果. 若安眠药无效, Z_i的差异仅由随机误差引起, 随机误差可认为服从正态分布$N(0,\sigma^2)$.故可假定$Z_1,\ldots,Z_n$为自$N(\mu,\sigma^2)$中抽取的简单随机样本, μ就是安眠药的平均效果. 安眠药是否有效, 就归结为检验如下假设

$$H_0:\ \mu=0\longleftrightarrow H_1:\ \mu\neq 0$$

因为$Z_1,\ldots,Z_n$被认为是来自正态总体$N(\mu,\sigma^2)$的简单样本, 故可用关于单个正态总体均值的t检验方法. 检验的否定域为

$$D=\{(Z_1,Z_2,\ldots,Z_n):|T|>t_{n-1}(\tfrac{\alpha}{2})\}$$

此处α为检验水平, $T=\sqrt{n}\,\bar{Z}/S$为检验统计量, 其中$\bar{Z}$和S^2为$Z_1,Z_2,\ldots,Z_n$ 的样本均值和样本方差.

例5.2.5 今有两台测量材料中某种金属含量的光谱仪A和B,为鉴定它们的质量有无显著差异, 对该金属含量不同的9件材料样品进行测量, 得到9对观察值如下:

$$\mu(\%):\quad 0.20,\ 0.30,\ 0.40,\ 0.50,\ 0.60,\ 0.70,\ 0.80,\ 0.90,\ 1.00.$$
$$\nu(\%):\quad 0.10,\ 0.21,\ 0.52,\ 0.32,\ 0.78,\ 0.59,\ 0.68,\ 0.77,\ 0.89.$$

问根据实验结果, 在$\alpha=0.01$下, 能否判断这两台光谱仪的质量有无显著差异?

解 将光谱仪A和B对9件样品的测定值记为$X_1,X_2,\ldots,X_9$,和$Y_1,Y_2,\ldots,Y_9$. 由于这9件样品金属含量不同, 因此$X_1,X_2,\ldots,X_9$ 不能看成来自同一总体. $Y_1,Y_2,\ldots,Y_9$也一样. 故需用成对比较. 记

$$Z_i=Y_i-X_i,\ i=1,2\ldots,9.$$

若这两光谱仪质量一样, 测量得到的每对数据的差异仅由随机误差引起. 随机误差可认为服从正态分布$N(0,\sigma^2)$.故可假定$Z_1,\ldots,Z_n$为自$N(\mu,\sigma^2)$中抽取的随机样本, 要检验

$$H_0:\ \mu=0\longleftrightarrow H_1:\ \mu\neq 0,\qquad \alpha=0.01$$

由表5.2.1可知此检验的否定域为

$$\{(Z_1,\cdots,Z_n):\ |T|>t_{n-1}(\alpha/2)\}$$

此处$n=9$.由题中数据算得:

$$\bar{Z}=\frac{1}{9}\sum_{i=1}^{9}Z_i=0.06,\quad S^2=\frac{1}{8}\sum_{i=1}^{9}(Z_i-\bar{Z})^2=0.01505,\quad S=0.12268$$

查表得 $t_{n-1}(\frac{\alpha}{2})=t_8(0.005)=3.3554$.由于

$$|T|=\left|\frac{\sqrt{n}\bar{Z}}{S}\right|=\frac{3\times 0.06}{0.12268}=1.47<3.3554$$

故无足够证据显示两台仪器有显著差异, 因此接受H_0.

四、两个正态总体方差比的检验

设$\mathbf{X}=(X_1,\dots,X_m)$是从正态总体$N(\mu_1,\sigma_1^2)$中抽取的简单随机样本,$\mathbf{Y}=(Y_1,\dots,Y_n)$是从正态总体$N(\mu_2,\sigma_2^2)$中抽取的简单随机样本, 且合样本$(\mathbf{X},\mathbf{Y})=(X_1,\dots,X_m;Y_1,\dots,Y_n)$独立. 讨论下列三类假设检验问题:

(10) $H_0:\ \sigma_2^2/\sigma_1^2=1\longleftrightarrow H_1:\ \sigma_2^2/\sigma_1^2\neq 1$;

(11) $H_0':\ \sigma_2^2/\sigma_1^2\le 1\longleftrightarrow H_1':\ \sigma_2^2/\sigma_1^2>1$;

(12) $H_0'':\ \sigma_2^2/\sigma_1^2\ge 1\longleftrightarrow H_1'':\ \sigma_2^2/\sigma_1^2<1$;

检验水平α给定.

记$\bar{X}$和S_1^2为$X_1,\dots,X_m$的样本均值和样本方差; $\bar{Y}$和S_2^2为$Y_1,\dots,Y_n$的样本均值和样本方差. 其中

$$S_1^2=\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\bar{X})^2,\qquad S_2^2=\frac{1}{n-1}\sum_{j=1}^{n}(Y_j-\bar{Y})^2.$$

下面将着重讨论μ_1和μ_2未知时方差比的检验方法,当μ_1和μ_2已知是方差比的检验如何处理, 我们将在后面给出一个说明.

首先讨论检验问题(1), 即

$$H_0:\ \sigma_2^2/\sigma_1^2\le 1\longleftrightarrow H_1:\ \sigma_2^2/\sigma_1^2>1$$

由于S_1^2和S_2^2分别是σ_1^2和σ_2^2的无偏估计, 并具有良好性质. 直观上看, S_2^2/S_1^2太小或者S_2^2/S_1^2太大时, H_0不象成立. 可设想否定域的形式为

$$\{(X_1,\dots,X_m;Y_1,\dots,Y_n):\ S_2^2/S_1^2<c_1\text{ 或 }S_2^2/S_1^2>c_2\},\quad c_1,\ c_2\text{ 待定}.$$

在$S_2^2/S_1^2=1$的条件下, 由推论2.4.4可知

$$F=S_2^2/S_1^2\sim F_{n-1,m-1}. \tag{2.12}$$

因此取检验统计量为$F=S_2^2/S_1^2$. 记$\theta=(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$,为了确定否定域中的临界值$c_1$, c_2,令

$$\alpha=P_\theta\left(S_2^2/S_1^2<c_1\text{ 或 }S_2^2/S_1^2>c_2|H_0\right)$$

满足上式要求的c_1和c_2有很多, 其中存在一对c_1,c_2最优, 但计算复杂, 使用不方便. 确定c_1,c_2的一个简单实用的方法是: 令

$$P_\theta\left(S_2^2/S_1^2<c_1|H_0\right)=\alpha/2,\quad P_\theta\left(S_2^2/S_1^2>c_2|H_0\right)=\alpha/2.$$

由上述两式和(2.12)易知临界值$c_1=F_{n-1,m-1}(1-\alpha/2)$, $c_2=F_{n-1,m-1}(\alpha/2)$, 所以检验问题(10)的水平为$\alpha$的接受域为

$$\bar{D}_{10} = \{(\mathbf{X}, \mathbf{Y}):\ F_{n-1,m-1}(1-\alpha/2) \leqslant S_2^2/S_1^2 \leqslant F_{n-1,m-1}(\alpha/2)\}$$

此接受域的表达式比否定域简单, 使用方便, 故此处采用接受域代替否定域.

用完全类似于(一) 中求检验问题(2)、(3)的方法可分别求得检验问题(11)和(12)的水平为α的否定域如下:

$$D_{11} = \{(X_1, \ldots, X_m; Y_1, \ldots, Y_n):\ S_2^2/S_1^2 > F_{n-1,m-1}(\alpha)\}$$
$$D_{12} = \{(X_1, \ldots, X_m; Y_1, \ldots, Y_n):\ S_2^2/S_1^2 < F_{n-1,m-1}(1-\alpha)\}$$

这里要注意一点的是: 当α是较小的数, 如$\alpha = 0.01$, 0.05时, 从$F-$分布的分位数表上查不到$F_{n-1,m-1}(1-\alpha)$的数值, 但利用第二章习题中已证明的事实

$$F_{n,\,m}(1-\alpha) = 1/F_{m,\,n}(\alpha), \tag{2.13}$$

使问题获得解决. 例如从表上查不到$F_{5,10}(1-0.01)$的值, 但可查到$F_{10,5}(0.01)$的值, 利用公式(2.13) 可知$F_{5,10}(1-0.01) = 1/F_{10,5}(0.01)$,从而可求得所要的数值.

注5.2.3 当μ_1和μ_2已知时, 方差比的检验方法简述如下: 当μ_1和μ_2已知时, σ_1^2和σ_2^2具有良好性质的无偏估计分别是

$$S_{1*}^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i-\mu_1)^2, \quad S_{2*}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu_2)^2$$

当$\sigma_1^2/\sigma_2^2 = 1$时, 利用推论2.4.1和$F$分布的定义, 容易证明

$$F_* = S_{2*}^2/S_{1*}^2 \sim F_{n,\,m}\,. \tag{2.14}$$

因此,取检验统计量F_*代替$F = S_2^2/S_1^2$, 完全类似于μ_1和μ_2未知情形的讨论, 可得到检验问题(10)-(12)的水平为α的否定域,只要注意在否定域中, 将确定临界值的F分布的自由度由$n-1, m-1$分别改为n, m即可. 详细结果见下列表5.2.4.

这种基于检验统计量服从F分布的检验方法, 称为F检验.

表5.2.4 两个正态总体方差比的假设检验

	H_0	H_1	检验统计量及其分布	否定域
μ_1	$\sigma_2^2 = \sigma_1^2$	$\sigma_2^2 \neq \sigma_1^2$	$F_* = s_{2*}^2/S_{1*}^2$	$F_* < F_{n,m}(1-\alpha/2)$
μ_2			$F_*\|_{\sigma_2^2=\sigma_1^2} \sim F_{n,m}$	或 $F_* > F_{n,m}(\alpha/2)$
已	$\sigma_2^2 \leq \sigma_1^2$	$\sigma_2^2 > \sigma_1^2$	$S_{1*}^2 = \frac{1}{m}\sum_{i=1}^m (X_i-\mu_1)^2$	$F_* > F_{n,m}(\alpha)$
知	$\sigma_2^2 \geq \sigma_1^2$	$\sigma_2^2 < \sigma_1^2$	$S_{2*}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i-\mu_2)^2$	$F_* < F_{n,m}(1-\alpha)$
μ_1	$\sigma_2^2 = \sigma_1^2$	$\sigma_2^2 \neq \sigma_1^2$	$F = s_2^2/S_1^2$	$F < F_{n-1,m-1}(1-\alpha/2)$
μ_2			$F\|_{\sigma_2^2=\sigma_1^2} \sim F_{n-1,m-1}$	或 $F > F_{n-1,m-1}(\alpha/2)$
未	$\sigma_2^2 \leq \sigma_1^2$	$\sigma_2^2 > \sigma_1^2$	$S_1^2 = \frac{1}{m-1}\sum_{i=1}^m (X_i-\bar{X})^2$	$F > F_{n-1,m-1}(\alpha)$
知	$\sigma_2^2 \geq \sigma_1^2$	$\sigma_2^2 < \sigma_1^2$	$S_2^2 = \frac{1}{n-1}\sum_{j=1}^n (Y_j-\bar{Y})^2$	$F < F_{n-1,m-1}(1-\alpha)$

例5.2.6 测得两批样本大小皆为6的电子器材电阻的均值$\bar{X} = 0.14, \bar{Y} = 0.139$, 样本标准差分别为$S_X = 0.0026$, $S_Y = 0.0024$,假设这两批器材的电阻分别服从$N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$, 均值方差皆未知, 且两组样本独立, 问这两批电子器件的电阻是否相同? $(\alpha = 0.05)$

解: 这个问题表面看是对两个正态总体均值差的检验,但我们不知道是否有$\sigma_1^2=\sigma_2^2$, 因此首先求两个正态总体方差是否相同的检验. 如果检验认为$\sigma_1^2=\sigma_2^2$, 然后再作两样本t检验. 如果经检验否定了$\sigma_1^2=\sigma_2^2$, 则我们不能用两样本的t检验方法检验均值差, 这就变成*Behrens-Fisher*问题, 将留在本节最后解决.

首先考虑下列检验问题:

$$(1)\quad H_0:\ \sigma_1^2=\sigma_2^2 \longleftrightarrow H_1:\ \sigma_1^2\neq\sigma_2^2,\ \ \alpha=0.05\,.$$

由表5.2.4可知此检验的接受域是

$$\{(\mathbf{X},\mathbf{Y}):\ F_{m-1,m-1}(1-\alpha/2)\le S_X^2/S_Y^2\le F_{m-1,n-1}(\alpha/2)\},$$

此处$m=n=6,\ S_X^2/S_Y^2=0.0026^2/0.0024^2=1.17$.由$\alpha=0.05$, 查F分布表得$F_{5,5}(0.025)=7.15$. 由于

$$1/7.15=F_{5,5}(1-0.025)<F=S_X^2/S_Y^2=1.17<F_{5,5}(0.025)=7.15,$$

故认为没有足够的证据否定H_0,因此接受H_0.

在接受上述检验后, 我们可以假定$\sigma_1^2=\sigma_2^2$,进一步考虑下列检验问题:

$$(2)\ \ H_0':\mu_1=\mu_2\longleftrightarrow H_1':\mu_1\neq\mu_2,\ \ \alpha=0.05\,.$$

由表5.2.3可知此检验的否定域为

$$\{(\mathbf{X},\mathbf{Y}):\ |T_w|>t_{n+m-2}(\alpha/2)\}$$

此处$m=n=6,\ \bar{X}=0.14,\ \bar{Y}=0.139,\ S_X=0.0026,\ S_y=0.0024$,因此有

$$S_w^2=\tfrac{1}{10}[5\times0.0026^2+5\times0.0024^2]=6.26\times10^{-6},\ \ S_w=0.0025.$$

由$\alpha=0.05$,查$t-$分布表得$t_{10}(0.0025)=2.228$. 由于

$$|T|=\sqrt{\tfrac{nm}{n+m}}\left|\tfrac{\bar{Y}-\bar{X}}{S_w}\right|=\sqrt{3}\times\left|\tfrac{0.14-0.139}{0.0025}\right|=0.6928<2.228$$

故没有充足的理由否定两批电子器件的电阻值相同,因此接受H_0'