数 理 统 计

上海财经大学　统计与管理学院

ConTenTs

第四章　假设检验

§ 4.1　假设检验的基本思想与概念

§ 4.2　正态总体参数假设检验

§ 4.3　其他分布参数的假设检验

§ 4.4　似然比检验与分布拟合检验

§ 4.5　正态性检验

§4.1 假设检验的基本思想与概念

假设检验问题:

在自然科学和社会科学等中，常常要对某些重要问题做出回答：是或否。

如月球比地球早形成吗？一种新药对某种病有效吗？某种股票会涨吗？等等。

为了回答这些问题，我们需要对感兴趣的问题进行试验或观察获得相关数据，根据这些数据决定是或否的过程称为**假设检验(hypothesis testing)**。

§4.1 假设检验的基本思想与概念

在总体X的分布完全未知，或只知其分布但不知其参数的情况下，我们对X的分布或分布中的参数作出某种假设，然后根据样本，用统计分析方法检验这一假设是否合理，从而作出接受或拒绝这一假设的决定.

假设检验
- 参数假设检验
- 非参数假设检验

总体分布已知，检验关于未知参数的某个假设

总体分布未知时的假设检验问题

§ 4.1 假设检验的基本思想与概念

假设检验:

- 参数检验
  - 单个总体均值、两总体均值的检验
  - 单个总体方差、两总体方差的检验
- 非参数检验
  - 分布拟合检验
  - 正态性检验
  - 游程检验、符号检验、秩和检验

§ 4.1 假设检验的基本思想与概念

4.1.1 假设检验问题

- 例4.1（女士品茶试验）一种奶茶由牛奶与茶按一定比例混合而成，可以先倒茶后倒奶（记为TM）,也可以反过来（记为MT）。某女士声称她可以鉴别是TM还是MT，周围品茶的人对此产生了议论，在场的费希尔（Fisher）也在思索这个问题，他提议做一项试验来检验如下假设（命题）是否可以接受。
- 假设H：该女士无此种鉴别能力。
- 他准备了10杯调制好的奶茶，TM与MT都有， Fisher记录下每一杯品尝的结果。

§4.1 假设检验的基本思想与概念

- 例4. 2 某厂生产的合金强度服从$N(\theta, 16)$,其中的设计值θ为不低于110(Pa)。为保证质量，该厂每天都要对生产情况做例行检查，以判断生 产是否正常进行,即该合金的平均强度不低于110(Pa)。某天从生产的产品中随机抽取25块合金，测得其强度值为$x_1, x_2, \ldots, x_{25}$,均值为$\bar{x} = 108.2$(Pa)，
- 问:当日生产是否正常?

4.1.1 假设检验问题

对这个实际问题可作如下分析：

(1) 是参数估计问题吗？

(2) 回答“是”还是“否”，假设检验问题。

(3) 命题“合金平均强度不低于110Pa”正确与否仅涉及如下两个参数集合：

$$\boldsymbol{\Theta}_0 = \{\boldsymbol{\theta}: \boldsymbol{\theta} \geq 110\} \quad \boldsymbol{\Theta}_1 = \{\boldsymbol{\theta}: \boldsymbol{\theta} < 110\}$$

这两个非空参数集合都称作统计假设，简称假设。

4.1.1 假设检验问题

(4) 我们的任务是利用所给总体$N(\theta, 16)$和样本均值$\bar{x} = 108.2$(Pa)去判断假设（命题）“$\theta \in \Theta_0$”是否成立。这里的“判断”在统计学中称为检验或检验法则。

(5)若假设可用一个参数的集合表示，该假设检验问题称为参数假设检验问题，否则称为非参数假设检验问题。例4. 2就是一个参数假设检验问题。

4.1.2 假设检验的基本步骤

一、建立假设

- 在假设检验中，常把一个被检验的假设称为原假设(null hypothesis)，用H_0表示，通常将不应轻易加以否定的假设作为原假设。当H_0被拒绝时而接收的假设称为对立假设或备择假设(alternative hypothesis)，用H_1表示，它们常常成对出现。

$$H_0:\ \theta \in \Theta_0 \qquad vs \qquad H_1: \theta \in \Theta_1$$

- 在例4. 2中，我们可建立如下两个假设：

$$H_0: \theta \geq 110 \qquad vs \qquad H_1: \theta < 110$$

4.1.2 假设检验的基本步骤

- 简单(simple)原假设，如 $H_0: \theta = \theta_0$.
  复杂或复合(composite)原假设，如 $H_0: \theta \geq 110$.

$H_0: \theta = \theta_0 \ vs \ H_1: \theta \neq \theta_0$ 双侧假设或双边假设

$H_0: \theta = \theta_0 \ vs \ H_1: \theta < \theta_0$ 单侧假设或单边假设

$H_0: \theta \geq \theta_0 \ vs \ H_1: \theta < \theta_0$ 单侧假设或单边假设

$H_0: \theta = \theta_0 \ vs \ H_1: \theta > \theta_0$ 单侧假设或单边假设

$H_0: \theta \leq \theta_0 \ vs \ H_1: \theta > \theta_0$ 单侧假设或单边假设

4.1.2 假设检验的基本步骤

二、选择检验统计量，给出拒绝域形式

- 检验准则：把样本空间划分成两个互不相交的部分W和$\overline{W}$，当样本属于W时，拒绝H_0；否则，就接受H_0. 我们称W为该检验的拒绝域(rejection region)， $\overline{W}$称为接受域(acceptance region)。
- 由样本对原假设进行判断总是通过一个统计量完成的，该统计量称为检验统计量(test statistic)。通过检验统计量可构造拒绝域W。

4.1.2 假设检验的基本步骤

在例4.2中，要检验的假设为

$$H_0: \theta \geq 110 \qquad\qquad H_1: \theta < 110$$

可考虑样本均值$\bar{X}$， $\bar{X}$ 愈大，意味着总体均值θ也愈大，因此，合理的拒绝域形如

$$W = \{(X_1, \dots, X_n): \bar{X} \leq c\} = \{\bar{X} \leq c\}$$

4.1.2 假设检验的基本步骤

三、选择显著性水平

检验可能犯以下两类错误：

- 第一类错误(Type I error)：H_0为真但观测值落在拒绝域中，从而拒绝原假设H_0。（拒真）

  拒真概率：第一类错误发生的概率，通常记为α。

$$\alpha = P(x \in W | H_0)$$

- 第二类错误(Type II error) : H_0不真（即H_1为真）但样本观测值落在接受域中，从而接受原假设H_0。（取伪）

  取伪概率：第二类错误发生的概率， 通常记为β。

$$\beta = P(x \in \overline{W} | H_1)$$

4.1.2 假设检验的基本步骤

表4. 1 检验的两类错误

观测数据情况	总体情况	
	H_0为真	H_1为真
$(x_1,\dots,x_n)\in W$	犯第一类错误	正确
$(x_1,\dots,x_n)\in \overline{W}$	正确	犯第二类错误

如何给出检验准则，使得两类错误都尽可能地小？
答案是做不到。

4.1.2 假设检验的基本步骤*

- 犯第一类错误的概率α和犯第二类错误的概率β可以用同一个函数表示，即所谓的**势函数或功效函数(power function)**。势函数是假设检验中最重要的概念之一。
- 定义4.1 设检验问题

$$H_0: \theta \in \Theta_0 \quad vs \quad H_1: \theta \in \Theta_1$$

的拒绝域为W，则样本观测值落在拒绝域内的概率称为该检验的势函数，记为：

$$g(\theta) = P_\theta(x \in W), \quad \theta \in \Theta = \Theta_0 \cup \Theta_1$$

4.1.2 假设检验的基本步骤*

- 势函数$g(\theta)$是定义在参数空间Θ上的一个函数。犯两类错误的概率都是参数θ的函数，并可由势函数算得，即：

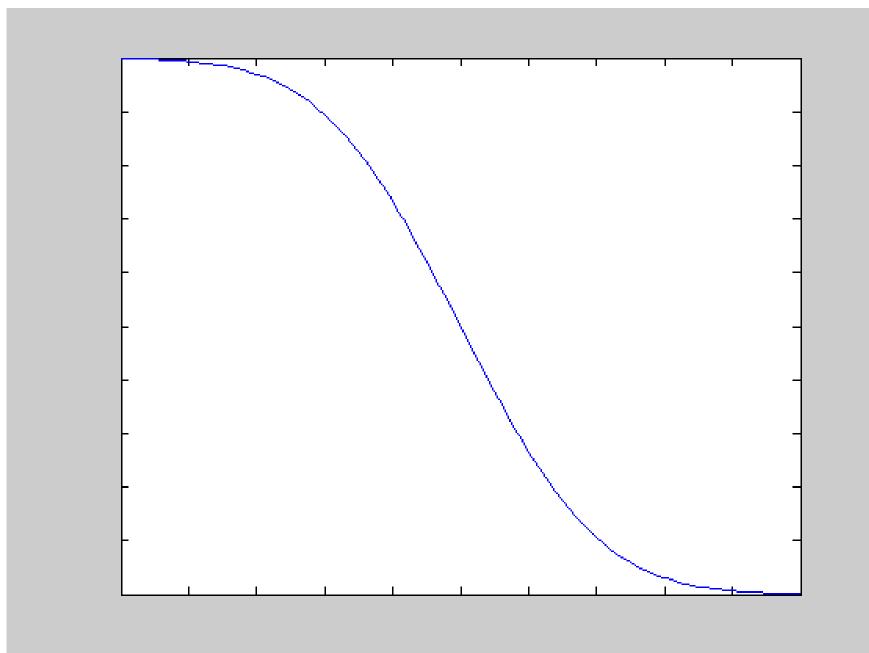
$$g(\theta)=\begin{cases}\alpha(\theta), & \theta\in\Theta_0\\ 1-\beta(\theta), & \theta\in\Theta_1\end{cases}$$

4.1.2 假设检验的基本步骤*

- 对例4. 2，其拒绝域为$W=\{\bar{X}\le c\}$，可以算出该检验的势函数

$$g(\theta)=P_\theta(\bar{X}\le c)=P_\theta\left(\frac{\bar{X}-\theta}{4/5}\le\frac{c-\theta}{4/5}\right)=\Phi\left(\frac{c-\theta}{4/5}\right)$$

这个势函数是θ的减函数 。



4.1.2 假设检验的基本步骤*

（续）利用这个势函数容易写出犯两类错误的概率分别为

$$\boldsymbol{\alpha(\theta)=\Phi\left(\frac{c-\theta}{4/5}\right),\qquad \theta\in\Theta_0}$$

和

$$\boldsymbol{\beta(\theta)=1-\Phi\left(\frac{c-\theta}{4/5}\right),\qquad \theta\in\Theta_1}$$

4.1.2 假设检验的基本步骤

由此可得如下结论：

- 当α减小时，c 也随之减小，必导致β 的增大；
- 当β减小时，c 会增大，必导致α的增大；

说明：在样本量一定的条件下不可能找到一个使α和β都小的检验。

英国统计学家 ***Neyman*** 和 ***Pearson*** 提出水平为α的显著性检验(significance test)的概念。

4.1.2 假设检验的基本步骤

- 定义4.2 对检验问题：

$$H_0: \theta \in \Theta_0 \quad vs \quad H_1: \theta \in \Theta_1$$

如果一个检验满足对任意的$\theta \in \Theta_0$，都有

$$g(\theta) \le \alpha,$$

则称该检验是显著性水平为 α 的显著性检验，

简称水平为α的检验(a level α test)。

4.1.2 假设检验的基本步骤

四、给出拒绝域

- 确定显著性水平后，可以定出检验的拒绝域W。
- 在例4.2中，若取$\alpha=0.05$, 由于$g(\theta)$关于θ单调减，只需要

$$g(110)=\Phi\left(\frac{5(c-110)}{4}\right)=0.05$$

成立即可。这给出c的值为

$$c=110+0.8u_{0.05}=110-0.8\times 1.645=108.684$$

检验的拒绝域为 $W=\{\bar{X}\le 108.684\}$

- 若令 $U=\frac{\bar{X}-110}{4/5}$,则拒绝域有另一种表示:

$$W=\{U\le u_{0.05}\}=\{U\le -1.645\}$$

4.1.2 假设检验的基本步骤

五、作出判断

- 在有了明确的拒绝域后，根据样本观测值我们可以做出判断：
  - 当$\bar{X} \leq 108.684$或$U \leq -1.645$时，则拒绝H_0，即接受H_1；
  - 当$\bar{X} > 108.684$或$U > -1.645$时，则接受H_0

在例4.2中，由于$\bar{x} = 108.2 < 108.684$

因此拒绝原假设，即认为该日生产不正常。

4.1.2 假设检验的基本步骤

❖ 假设检验的步骤总结

1. 建立假设 $H_0 \quad vs \quad H_1$;

2. 选取检验统计量 $T(X_1, \dots, X_n)$, 使得当H_0成立时, T 的分布完全已知，并确定拒绝域W的形状;

3. 在给定的显著性水平α下，确定拒绝域 W;

4. 根据样本观测值计算检验统计量$T(X_1, \dots, X_n)$, 判断其是否属于拒绝域 W，做出最终判断。

4.1.3 检验的p值

- 假设检验的结论通常是简单的: 在给定的显著性水平下，不是拒绝原假设就是保留原假设。然而有时也会出现这样的情况：在一个较大的显著性水平（α=0.05）下得到拒绝原假设的结论，而在一个较小的显著水平（α=0.01）下却会得到相反的结论。
- 这种情况在理论上很容易解释：因为显著性水平变小后会导致检验的拒绝域变小，于是原来落在拒绝域中的观测值就可能落入接受域。
- 但这种情况在应用中会带来一些麻烦：假如这时一个人主张选择显著性水平α=0.05，而另一个人主张选α=0.01，则第一个人的结论是拒绝H_0，而后一个人的结论是接受H_0，我们该如何处理这一问题呢？

4.1.3 检验的p值

- 下面用例4.2来进行讨论。
- 例4.2的拒绝域$W=\{\bar{X}\le 110+0.8u_\alpha\}$,或表示为$W=\{U\le u_\alpha\}$,其中$U=1.25(\bar{x}-110)=-2.25$.对给定的显著性水平，表4.2列出了相应的拒绝域和检验结论。

表4.2 例4.2中的拒绝域

显著性水平	拒绝域	对应的结论（$u=-2.25$）
$\alpha=0.1$	$U\le -1.282$	拒绝H_0
$\alpha=0.05$	$U\le -1.645$	拒绝H_0
$\alpha=0.025$	$U\le -1.96$	拒绝H_0
$\alpha=0.01$	$U\le -2.326$	接受H_0
$\alpha=0.005$	$U\le -2.576$	接受H_0

我们看到，不同的α有不同的结论。

4.1.3 检验的p值

- 现在换一个角度来看，在$\theta=110$时，U的分布是$N(0,1)$。此时可算得，$P(U\le-2.25)=0.0122$，若以0.0122为基准来看上述检验问题，可得
  - 当α<0.0122时，$u_\alpha<-2.25$。于是-2.25就不在$\{U\le u_\alpha\}$中，此时应接受原假设H_0;
  - 当α≥0.0122时，$u_\alpha\ge-2.25$。于是-2.25就落在$\{U\le u_\alpha\}$中，此时应拒绝H_0。

由此可以看出，0.0122是能用观测值-2.25做出“拒绝H_0”的最小的显著性水平，这就是p值。

4.1.3 检验的p值

- 定义4.3 在一个假设检验问题中，利用观测值能够做出拒绝原假设的最小显著性水平称为检验的p 值。

- 引进检验的p 值的概念有明显的好处:
  - 第一，它比较客观，避免了事先确定显著性水平;
  - 其次，由检验的p 值与人们心目中的显著性水平α进行比较可以很容易作出检验的结论:

    如果$\alpha \geq p$，则在显著性水平α下拒绝 H_0;

    如果$\alpha < p$，则在显著性水平α下接受 H_0.
  - p 值在应用中很方便，如今的统计软件中对检验问题一般都会给出检验的p 值。

§ 4.2 正态总体参数假设检验

4.2.1 单个正态总体均值的检验

- 参数假设检验常见的有三种基本形式

（1）$H_0: \mu \le \mu_0 \quad vs \quad H_1: \mu > \mu_0$
（2）$H_0: \mu \ge \mu_0 \quad vs \quad H_1: \mu < \mu_0$
（3）$H_0: \mu = \mu_0 \quad vs \quad H_1: \mu \ne \mu_0$

- 当备择假设H_1在原假设H_0一侧时的检验称为**单侧检验(one-sided test)**；
- 当备择假设H_1分散在原假设H_0两侧时的检验称为**双侧检验(two-sided test)**。

4.2.1 单个正态总体均值的检验

一、已知σ时的U检验

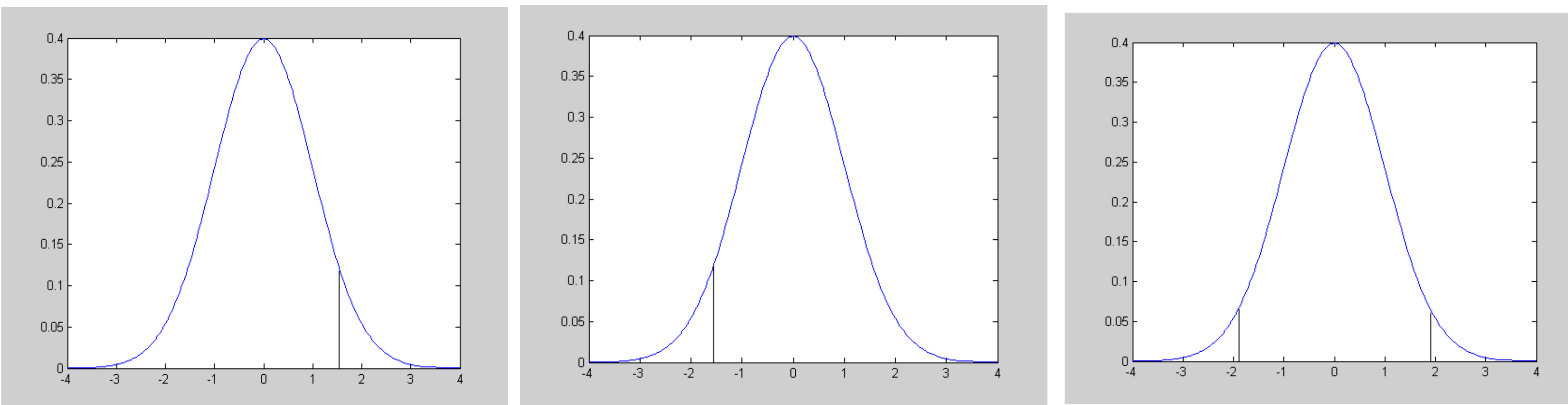
设$X_1, \dots, X_n$是来自$N(\mu, \sigma^2)$的样本，考虑关于μ的检验问题。检验统计量可选为

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

当$\mu = \mu_0$时， $U \sim N(0,1)$. 该检验用的检验统计量是U统计量，故一般称为U检验。

三种假设的拒绝域形式分别见下图：

4.2.1 单个正态总体均值的检验



$(a)\ H_1: \mu > \mu_0$　　$(b)\ H_1: \mu < \mu_0$　　$(c)\ H_1: \mu \neq \mu_0$

$W = \{U \geq u_{1-\alpha}\}$　　$W = \{U \leq -u_{1-\alpha}\}$　　$W = \{|U| \geq u_{1-\alpha/2}\}$

图4.1　　U检验的拒绝域

4.2.1 单个正态总体均值的检验*

- 下面以$H_0: \mu \le \mu_0 \quad vs \quad H_1: \mu > \mu_0$为例说明：

  由$P_{\mu_0}(u \ge c) = \alpha$可推出具体的拒绝域为

$$W = \{U \ge u_{1-\alpha}\}$$

  该检验的势函数是U的函数，它可用正态分布写出，具体为：

$$g(\mu) = 1 - \Phi\left(\frac{\sqrt{n}(\mu_0-\mu)}{\sigma} + u_{1-\alpha}\right)$$

4.2.1 单个正态总体均值的检验*

势函数是μ的增函数（见图4.2），只要$g(\mu_0)=\alpha$就可保证在$\mu\le\mu_0$时有$g(\mu)\le\alpha$。

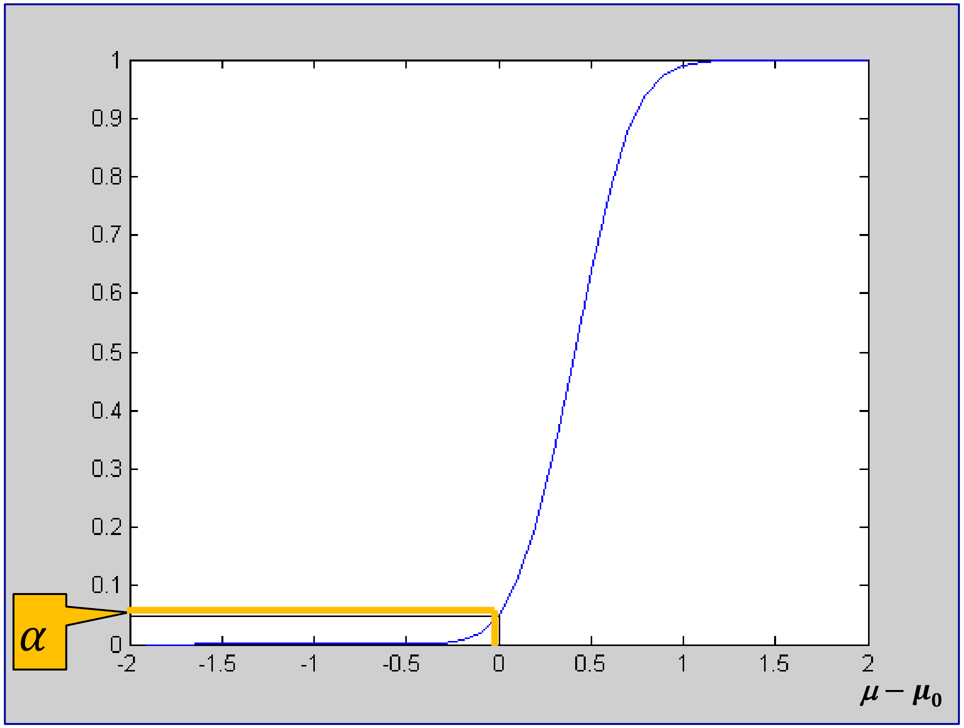


图4.2 情形(a)下$g(\mu-\mu_0)$的图形

4.2.1 单个正态总体均值的检验*

- 对单侧检验$H_0: \mu \geq \mu_0 \quad vs \quad H_1: \mu < \mu_0$是类似的，只是拒绝域变为: $W = \{U \leq u_\alpha\}$

  其势函数为： $g(\mu) = \Phi\left(\frac{\sqrt{n}(\mu_0-\mu)}{\sigma} + u_\alpha\right)$

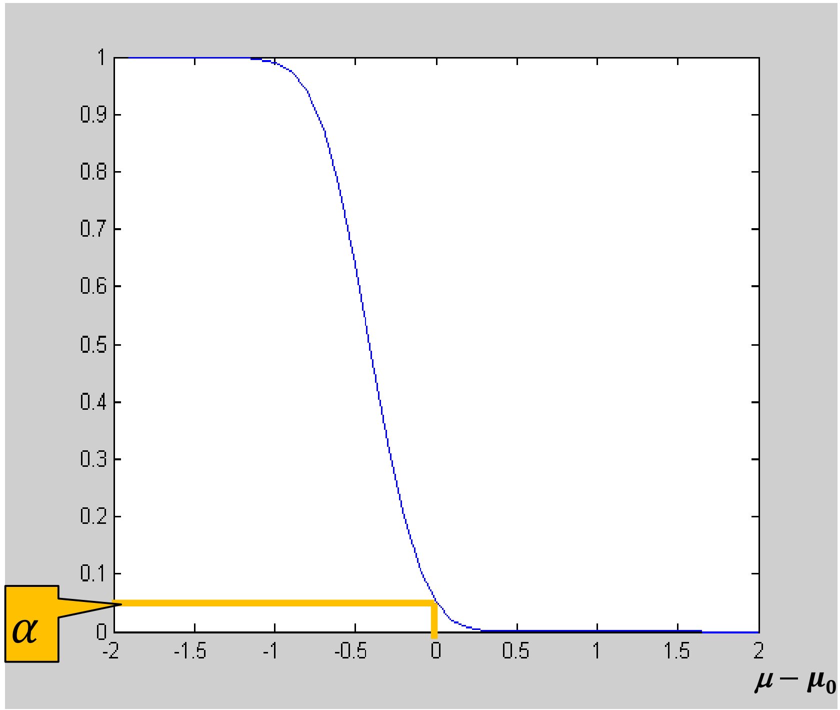
- 对双侧检验问题 $H_0: \mu = \mu_0 \; vs \; H_1: \mu \neq \mu_0$，拒绝域为

$$W = \left\{|U| \geq u_{1-\alpha/2}\right\}$$

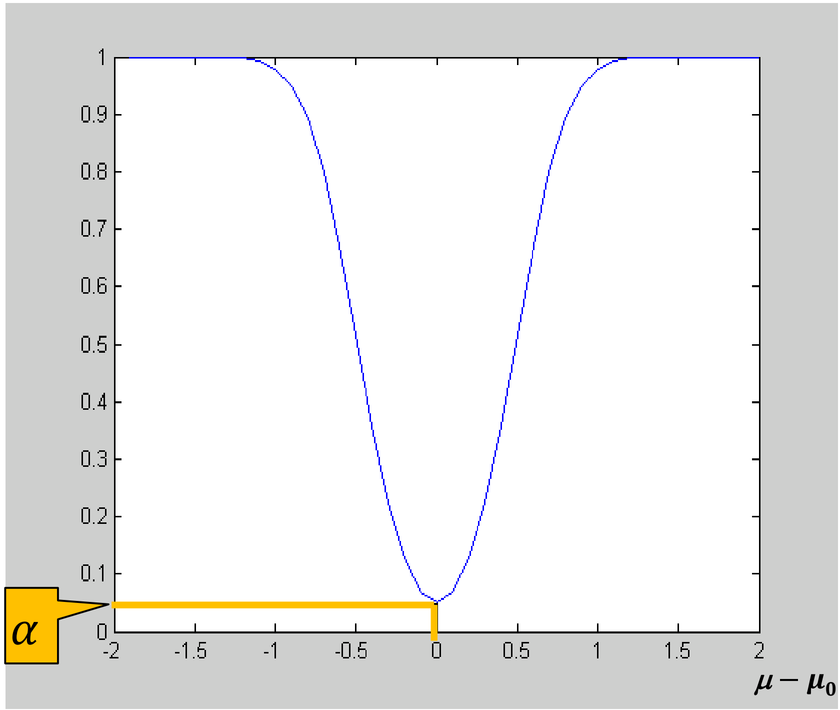
其势函数为:

$$g(\mu) = 1 - \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} + u_{1-\alpha/2}\right) + \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} + u_{\alpha/2}\right)$$

4.2.1 单个正态总体均值的检验*



(b)



(c)

图4.3　情形 $(b)(c)$下$g(\mu-\mu_0)$的图形

4.2.1 单个正态总体均值的检验

- 例4.3某公司主管称他们的产品每箱不合格数不超过15个。为了验证该声称是否属实，抽取了36箱做调查，发现平均每箱不合格数为17。已知每箱的不合格数服从正态分布，且方差为9。问：在0.05的显著性水平下，能否验证该公司主管的声称存在问题？

4.2.1 单个正态总体均值的检验

- **解:** 这是一个假设检验的问题，总体$X \sim N(\mu, 9)$,

检验假设: $H_0: \mu \le 15 \quad v.s. \quad H_1: \mu > 15$

取检验统计量为$U = \dfrac{\bar{X}-15}{\sigma/\sqrt{n}}$,

拒绝域为：$\{U \ge u_{1-\alpha}\}$.

取显著性水平α=0.05，查表知 $u_{0.95}$=1.65。

用观测值可计算得:

$$\bar{x} = 17, \quad U = \frac{\sqrt{36}(17-15)}{3} = 4 > 1.65$$

U值落入拒绝域内，拒绝原假设，认为公司主管的声称不属实。

该检验的p值为:

$$p - value = P(U > 4) \approx 0.000$$

4.2.1 单个正态总体均值的检验

二、σ未知时的 T 检验

由于σ未知，一个自然的想法是将$U=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$中未知的$\sigma$替换成样本标准差 S，这就形成 T 检验统计量：

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

三种假设的检验拒绝域分别为：

$$\{T\geq t_{1-\alpha}(n-1)\},$$

$$\{T\leq t_{\alpha}(n-1)\},$$

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

4.2.1 单个正态总体均值的检验

- 例4.4 某厂生产的某种铝材的长度服从正态分布，其均值设定为240厘米。现从该厂抽取5件产品，测得其长度为（单位：厘米）

239.7 239.6 239 240 239.2

试判断该厂此类铝材的长度是否满足设定要求?

4.2.1 单个正态总体均值的检验

- 解：假设：$H_0: \mu = 240 \quad vs. \quad H_1: \mu \neq 240$

采用T检验，检验统计量 $T = \frac{\sqrt{n}(\bar{X}-240)}{S}$，

拒绝域为: $\{|T| \geq t_{1-\alpha/2}(n-1)\}$

若取α=0.05，则 $t_{0.975}(4)$= 2.776.

现由样本计算得到:$\bar{x} = 239.5, s = 0.4,$

故：$T = \sqrt{5} \cdot |239.5 - 240|/0.4 = 2.7951$

由于2.7951>2.776，故拒绝原假设，认为该厂生产的铝材的长度不满足设定要求。

该检验的p值为：

$$p - value = P(|T| > 2.7951) = 0.0491$$

4.2.1 单个正态总体均值的检验

- 例4.5 一个备择假设为$H_1: \mu > 150.0$的假设检验问题。显著性水平是0.01，样本量为4，样本均值为156.3，样本标准差为15.4，并且有足够的理由相信总体服从正态分布。试对该假设检验问题进行分析。

4.2.1 单个正态总体均值的检验

- **解：** 这是一个关于正态均值的单侧假设检验问题。

$$H_0:\ \mu \le 150.0,\quad H_1:\ \mu > 150.0$$

采用T 检验，检验统计量 $T = \dfrac{\sqrt{n}(\bar{X}-150.0)}{S}$，

拒绝域为: $\{T \ge t_{1-\alpha}(n-1)\}$

若取α =0.01，则 $t_{0.99}(3)$= 4.541.

现由样本计算得到: $\bar{x} = 156.3, s = 15.4$，故:

$$T = \sqrt{4} \cdot (156.3 - 150.0)/15.4 \approx 0.8182$$

由于0.8182<4.541，故不拒绝原假设，认为 μ 不大于150.0。

该检验的p值为:

$$p - value = P(T > 0.82) = 0.2362$$

4.2.1 单个正态总体均值的检验

表4.3 单个正态总体均值的假设检验

检验法	条件	原假设H_0	备择假设H_1	检验统计量	拒绝域	p值
U检验	σ已知	$\mu \le \mu_0$ $\mu \ge \mu_0$ $\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \ne \mu_0$	$U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$\{U \ge u_{1-\alpha}\}$ $\{U \le u_{\alpha}\}$ $\{\lvert U\rvert \ge u_{1-\alpha/2}\}$	$1 - \Phi(u_0)$ $\Phi(u_0)$ $2(1 - \Phi(\lvert u_0\rvert))$
T检验	σ未知	$\mu \le \mu_0$ $\mu \ge \mu_0$ $\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \ne \mu_0$	$T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$	$\{T \ge t_{1-\alpha}(n-1)\}$ $\{T \le t_{\alpha}(n-1)\}$ $\{\lvert T\rvert \ge t_{1-\alpha/2}(n-1)\}$	$P(T \ge t_0)$ $p(T \le t_0)$ $P(\lvert T\rvert \ge \lvert t_0\rvert)$

4.2.2 假设检验与置信区间的关系

- 这里用的检验统计量与3.6.3节中置信区间所用的枢轴量是相似的。这不是偶然的，两者之间存在非常密切的关系。
- 设$X_1, \ldots, X_n$是来自正态总体$N(\mu, \sigma^2)$的样本，现在σ未知场合讨论关于均值μ的检验问题。

  考虑双侧检验问题:

$$H_0: \mu = \mu_0 \quad vs \quad H_1: \mu \neq \mu_0$$

检验统计量 $T = \dfrac{\sqrt{n}(\bar{X}-\mu_0)}{S}$,

拒绝域为: $\{|T| \geq T_{1-\alpha/2}(n-1)\}$

4.2.2 假设检验与置信区间的关系

则水平为α的检验接受域为

$$\overline{W} = \left\{ |\bar{X} - \mu_0| \le \frac{S}{\sqrt{n}} T_{1-\alpha/2}(n-1) \right\}$$

它可以改写为

$$\overline{W} = \left\{ \bar{X} - \frac{S}{\sqrt{n}} T_{1-\alpha/2}(n-1) \le \mu_0 \le \bar{X} + \frac{S}{\sqrt{n}} T_{1-\alpha/2}(n-1) \right\}$$

并且有$P_{\mu_0}(\overline{W}) = 1 - \alpha$,这里$\mu_0$并无限制.

若让μ_0 在$(-\infty, \infty)$内取值，就可得到μ 的$1-\alpha$ 置信区间：

$$\bar{X} \pm \frac{S}{\sqrt{n}} \mathrm{t}_{1-\alpha/2}(n-1)$$

4.2.2 假设检验与置信区间的关系

- 反之若有一个如上的1-α置信区间，也可获得关于$H_0: \mu = \mu_0$的水平为α的显著性检验。
  所以: “正态均值μ的1-α置信区间”与“关于$H_0: \mu = \mu_0$的双侧检验问题的水平α的检验”是一一对应的。

- 类似地，“参数μ的1-α置信下限”与“关于$H_0: \mu \le \mu_0$的单侧检验问题的水平α的检验”是 一一对应的。
  参数μ的1-α置信上限与另一个单侧检验也是 一一对应的。

4.2.3 两个正态总体均值差的检验

- 设$X_1,\dots,X_m$是来自正态总体$N(\mu_1,\sigma_1^2)$的样本，$Y_1,\dots,Y_n$是来自另一个正态总体$N(\mu_2,\sigma_2^2)$的样本，两个样本相互独立。考虑如下三类检验问题：

$$
\begin{aligned}
&(1) H_0: \mu_1-\mu_2 \le 0 \quad vs \quad H_1: \mu_1-\mu_2 > 0\\
&(2) H_0: \mu_1-\mu_2 \ge 0 \quad vs \quad H_1: \mu_1-\mu_2 < 0\\
&(3) H_0: \mu_1-\mu_2 = 0 \quad vs \quad H_1: \mu_1-\mu_2 \ne 0
\end{aligned}
$$

这里对常用的两种情形进行讨论。

4.2.3 两个正态总体均值差的检验

一、σ_1,σ_2已知时的两样本U 检验

此时$\mu_1-\mu_2$的点估计$\bar{X}-\bar{Y}$的分布完全已知，

$$\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2,\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n})$$

由此可采用U检验方法，检验统计量为

$$U=(\bar{X}-\bar{Y})\Big/\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$$

当$\mu_1=\mu_2$时， $U\sim N(0,1)$.

4.2.3 两个正态总体均值差的检验

检验的拒绝域取决于备择假设的具体内容。

- 对检验问题（1），检验的拒绝域与p值为

$$W=\{U\geq u_{1-\alpha}\},\qquad p=1-\Phi(u_0)$$

其中$u_0=(\bar{x}-\bar{y})\Big/\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$是由样本计算得到的检验统计量的值。

- 对检验问题（2），检验的拒绝域与p值为

$$W=\{U\leq u_{\alpha}\},\qquad p=\Phi(u_0)$$

- 对检验问题（3），检验的拒绝域与p值为

$$W=\left\{|U|\geq u_{1-\alpha/2}\right\},\qquad p=2(1-\Phi(|u_0|))$$

4.2.3 两个正态总体均值差的检验

二、$\sigma_1=\sigma_2=\sigma$但未知时的两样本T检验

T检验统计量：$T=\dfrac{(\bar{X}-\bar{Y})}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$,

其中$S_w^2=((m-1)S_X^2+(n-1)S_Y^2)/(m+n-2)$

三种假设的检验拒绝域分别为：

$$\{T\geq t_{1-\alpha}(m+n-2)\},$$

$$\{T\leq t_{\alpha}(m+n-2)\},$$

$$\left\{|T|\geq t_{1-\frac{\alpha}{2}}(m+n-2)\right\}.$$

4.2.3 两个正态总体均值差的检验

表4.4 两个正态总体均值差的检验

检验法	条件	原假设 H_0	备择假设H_1	检验统计量	拒绝域	p值
U检验	σ_1,σ_2 已知	$\mu_1 \le \mu_2$ $\mu_1 \ge \mu_2$ $\mu_1 = \mu_2$	$\mu_1 > \mu_2$ $\mu_1 < \mu_2$ $\mu_1 \ne \mu_2$	$U_1 = \dfrac{(\bar{X}-\bar{Y})}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}}$	$\{U_1 \ge u_{1-\alpha}\}$ $\{U_1 \le u_\alpha\}$ $\{\lvert U_1\rvert \ge u_{1-\alpha/2}\}$	$1-\Phi(u_1)$ $\Phi(u_1)$ $2(1-\Phi(\lvert u_1\rvert))$
T检验	σ_1,σ_2 未知 $\sigma_1=\sigma_2$	$\mu_1 \le \mu_2$ $\mu_1 \ge \mu_2$ $\mu_1 = \mu_2$	$\mu_1 > \mu_2$ $\mu_1 < \mu_2$ $\mu_1 \ne \mu_2$	$T_1 = \dfrac{(\bar{X}-\bar{Y})}{S_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$	$\{T \ge t_{1-\alpha}(\mathrm{m+n-2})\}$ $\{T \le t_\alpha(\mathrm{m+n-2})\}$ $\{\lvert T\rvert \ge t_{1-\alpha/2}(\mathrm{m+n-2})\}$	$P(T_1 \ge t_1)$ $p(T_1 \le t_1)$ $P(\lvert T_1\rvert \ge \lvert t_1\rvert)$

注：$S_w^2 = ((m-1)S_X^2 + (n-1)S_Y^2)/(m+n-2)$

4.2.3 两个正态总体均值差的检验

续表4.4 两个正态总体均值差的检验

检验法	条件	原假设 H_0	备择假设 H_1	检验统计量	拒绝域	P值
大样本U检验	σ_1,σ_2 未知 m,n 充分大	$\mu_1 \le \mu_2$ $\mu_1 \ge \mu_2$ $\mu_1 = \mu_2$	$\mu_1 > \mu_2$ $\mu_1 < \mu_2$ $\mu_1 \ne \mu_2$	$U_2 = \dfrac{(\bar{X}-\bar{Y})}{\sqrt{\dfrac{S_X^2}{m}+\dfrac{S_Y^2}{n}}}$	$\{U_2 \ge u_{1-\alpha}\}$ $\{U_2 \le u_\alpha\}$ $\{\lvert U_2\rvert \ge u_{1-\alpha/2}\}$	$1-\Phi(u_2)$ $\Phi(u_2)$ $2(1-\Phi(\lvert u_2\rvert))$
近似T检验	σ_1,σ_2 未知 m,n 不很大	$\mu_1 \le \mu_2$ $\mu_1 \ge \mu_2$ $\mu_1 = \mu_2$	$\mu_1 > \mu_2$ $\mu_1 < \mu_2$ $\mu_1 \ne \mu_2$	$T_2 = \dfrac{(\bar{X}-\bar{Y})}{\sqrt{\dfrac{S_X^2}{m}+\dfrac{S_Y^2}{n}}}$	$\{T_2 \ge t_{1-\alpha}(l)\}$ $\{T_2 \le t_\alpha(l)\}$ $\lvert T_2\rvert \ge t_{1-\alpha/2}(l)\}$	$P(T_2 \ge t_2)$ $p(T_2 \le t_2)$ $P(\lvert T_2\rvert \ge \lvert t_2\rvert)$

注： $l = S_0^4 \Big/ \left(\dfrac{S_X^4}{m^2(m-1)} + \dfrac{S_Y^4}{n^2(n-1)}\right)$

4.2.3 两个正态总体均值差的检验

- 例4.6 某厂铸造车间为提高铸件的耐磨性而试制了一种镍合金铸件以取代铜合金铸件，为此，从两种铸件中各抽取一个容量分别为8和9的样本，测得其硬度为

镍合金：76.43 76.21 73.58 69.69
65.29 70.83 82.75 72.34

铜合金：73.66 64.27 69.34 71.37
69.77 68.12 67.27 68.07 62.61

根据经验，硬度服从正态分布，且方差保持不变。试在显著性水平$\alpha=0.05$下判断镍合金的硬度是否有明显提高。

4.2.3 两个正态总体均值差的检验

- **解**：用X表示镍合金的硬度，Y表示铜合金的硬度，

则由假定，$X\sim N(\mu_1,\sigma^2), Y\sim N(\mu_2,\sigma^2)$.

要检验的假设是:$H_0: \mu_1=\mu_2 \quad vs \quad H_1: \mu_1>\mu_2$

检验统计量$T=\frac{(\bar{X}-\bar{Y})}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$，$S_w^2=((m-1)S_X^2+(n-1)S_Y^2)/(m+n-2)$

拒绝域为$\{T\geq t_{1-\alpha}(m+n-2)\}$.

经计算，$\bar{x}=73.39, \bar{y}=68.2756$, $(m-1)S_X^2=\sum_{i=1}^{8}(x_i-\bar{x})^2=205.7958$, $(n-1)S_Y^2\ \sum_{i=1}^{9}(y_i-\bar{y})^2=91.1552$, 从而

$$s_w=\sqrt{\frac{1}{8+9-2}(205.7958+91.1552)}=4.4494$$

4.2.3 两个正态总体均值差的检验

$$T=\frac{73.39-68.2756}{4.4494\cdot\sqrt{\frac{1}{8}+\frac{1}{9}}}=2.3656$$

查表知，$t_{0.95}(15)=1.7531$,

由于 $t>t_{0.95}(15)$,故拒绝原假设，可判断镍合金硬度有显著提高。

该检验的p值为：

$$P-value=P(T>2.3656)=0.0159$$

4.2.3 两个正态总体均值差的检验

- 例4.7 一个备择假设为$\mu_1 - \mu_2 > 100$的假设检验问题，显著性水平为0.01，样本1的个数为4，均值为1356.3，标准差为125.4；样本2的个数为6，均值为1168.9，标准差为123.7。有足够的理由相信总体均服从正态分布，且知道 $\sigma_1^2 = \sigma_2^2$ 。试对该假设检验问题进行分析。

4.2.3 两个正态总体均值差的检验

- **解**：用X表示样本1，Y表示样本2，则由假定，$X\sim N(\mu_1,\sigma^2), Y\sim N(\mu_2,\sigma^2)$. 要检验的假设是:

$$H_0: \mu_1-\mu_2\le 100 \quad vs \quad H_1: \mu_1-\mu_2>100$$

检验统计量$T=\frac{(\bar{X}-\bar{Y})-100}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$， $S_w^2=((m-1)S_X^2+(n-1)S_Y^2)/(m+n-2)$

拒绝域为$\{T\ge t_{1-\alpha}(m+n-2)\}$.

计算得：$\bar{x}=1356.3, \bar{y}=1168, (m-1)S_X^2=47175.48$，

$(n-1)S_Y^2=76508.45$， 从而

$$s_w=\sqrt{\frac{1}{4+6-2}\left(47175.48+76508.45\right)}\approx 124.3402$$

4.2.3 两个正态总体均值差的检验

$$T=\frac{(1356.3-1168.9)-100}{124.3402\cdot\sqrt{\frac{1}{4}+\frac{1}{6}}}\approx 1.089$$

查表知，$t_{0.99}(8)=2.8965$,

由于 $T<t_{0.99}(8)$,故不拒绝原假设，不能认$\mu_1-\mu_2>100$。

该检验的p值为：

$$P-value=P(T>1.089)=0.1539$$

4.2.3 两个正态总体均值差的检验

- 例4.8为了比较两种油漆，一个顾客检验服务中心分别抽取了4桶一加仑的A品牌油漆和4桶一加仑的B品牌油漆。涂刷结果显示：每桶A品牌油漆平均可涂刷546平方英尺，标准差为31平方英尺。而每桶B品牌油漆平均可涂刷492平方英尺，标准差为26平方英尺。假设两种油漆涂刷面积的总体均服从正态分布，且具有相同方差。根据样本数据能否认为每桶A品牌油漆平均涂刷面积大于B品牌？以0.05的水平进行假设检验。

4.2.3 两个正态总体均值差的检验

- **解：** 用X表示每桶品牌A油漆涂刷面积，Y表示每桶品牌B油漆涂刷面积，则由假定：

$$X \sim N(\mu_1, \sigma^2), Y \sim N(\mu_2, \sigma^2)$$

要检验的假设是:

$$H_0: \mu_1 - \mu_2 \le 0 \quad vs \quad H_1: \mu_1 - \mu_2 > 0$$

检验统计量$T = \frac{(\bar{X}-\bar{Y})}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$, $S_w^2 = ((m-1)S_X^2 + (n-1)S_Y^2)/(m+n-2)$

拒绝域为$\{T \ge t_{1-\alpha}(m+n-2)\}$.

由：$\bar{x} = 546, \bar{y} = 49, (m-1)S_X^2 = \sum_{i=1}^{4}(x_i - \bar{x})^2 = 2883, (n-1)S_Y^2 = \sum_{i=1}^{4}(y_i - \bar{y})^2 = 2028,$

从而 $s_w = \sqrt{\frac{1}{4+4-2}\left(2883 + 2028\right)} \approx 28.609$

4.2.3 两个正态总体均值差的检验

$$T=\frac{546-492}{28.609\cdot\sqrt{\frac{1}{4}+\frac{1}{4}}}\approx 2.67$$

查表知，$T_{0.95}(6)=1.943$,

由于 $T>t_{0.95}(6)$,故拒绝原假设，认为$\mu_1-\mu_2>0$。
该检验的p值为：

$$P-value=P(T>2.67)=0.0185$$

4.2.4 成对数据检验

- 例4.9 为了比较两种谷物种子的优劣，特选取10块土质不全相同的土地，并将每块土地分为面积相同的两部分，分别种植这两种种子，施肥与田间管理在20块小块土地上都是一样，下面是各小块的单位产量：

土地	1	2	3	4	5	6	7	8	9	10
种子一的单位产量 x	23	35	29	42	39	29	37	34	35	28
种子二的单位产量 y	30	39	35	40	38	34	36	33	41	31
差$\boldsymbol{d}=\boldsymbol{x}-\boldsymbol{y}$	-7	-4	-6	2	1	-5	1	1	-6	-3

假定单位产量服从正态分布，试问：两种种子的平均产量在显著性水平$\alpha=0.05$上有无显著差异？

4.2.4 成对数据检验

❖ **解：** 假定$X\sim N(\mu_1,\sigma_1^2), Y\sim N(\mu_2,\sigma_2^2)$, 且$X$与$Y$独立，这里假定两个总体的方差相等是合理的。我们先用二样本T检验讨论此问题。记两种种子的单位产量的样本均值分别为$\bar{X},\bar{Y}$,样本方差分别为S_X^2, S_Y^2。对如下检验问题：

$$H_0: \mu_1 = \mu_2 \quad vs \quad H_1: \mu_1 \neq \mu_2$$

作出判断。在假定$\sigma_1^2 = \sigma_2^2 = \sigma^2$下，采用二样本$T$检验，检验统计量$T_1$与拒绝域$W_1$分别是：

4.2.4 成对数据检验

$$T_1=\frac{\bar{X}-\bar{Y}}{S_w/\sqrt{\frac{n}{2}}}$$

$$W_1=\{|T_1|>t_{1-\frac{\alpha}{2}}(2n-2)\}$$

其中$S_w^2=(S_x^2+S_y^2)/2$，α是给定的显著性水平。

由给出的 数据可算得：

$$\bar{x}=33.1\quad \bar{y}=35.7$$

$$s_x^2=33.2110\quad s_y^2=14.2333\quad s_w^2=23.7222$$

4.2.4 成对数据检验

从而可算得两样本的T检验统计量的值

$$s_w = \sqrt{23.7222} = 4.8705,$$

$$T_1 = \frac{33.1 - 35.7}{\dfrac{4.8705}{\sqrt{10/2}}} = -1.1937.$$

若给定$\alpha = 0.05$，查表得$t_{0.975}(18) = 2.1009$，由于$|T_1| < 2.1009$，故不应拒绝原假设，即认为两种种子的单位产量平均值没有显著差别。此处检验的p值为0.2467。

4.2.4 成对数据检验

下面我们换一个角度来谈此问题。在这个问题中出现了成对数据，同一块土地上用两种种子得两个产量，其差$D_i = X_i - Y_i (i = 1, 2, \dots, 10)$排除了土质差异这个不可控因素的影响，主要反映两种种子的优劣。对这种信息我们应加以利用。在正态性假定下，$D = X - Y \sim N(\mu, \sigma_D^2)$，其中$\mu = \mu_1 - \mu_2$. 原先要比较$\mu_1$与$\mu_2$的大小，如今则转化为考察$\mu$是否为零，即考察如下检验问题：

4.2.4 成对数据检验

$$H_0: \mu = 0 \quad vs \quad H_1: \mu \neq 0$$

即把双样本的检验问题转化为单样本T检验问题，这时检验的T统计量为：

$$T_2 = \bar{D}/(S_D/\sqrt{n})$$

其中 $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$， $S_D = \left(\frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2\right)^{1/2}$

在给定的显著性水平α下，该检验问题的拒绝域是

$$W = \{|T_2| \geq t_{1-\frac{\alpha}{2}}(n-1)\}$$

这就是成对数据的T检验。

4.2.4 成对数据检验

在本例中可算得

$$n = 10 \quad \bar{d} = -2.6 \quad S_d = 3.5024$$

于是 $T_2 = \dfrac{-2.6}{3.5024/\sqrt{10}} = \dfrac{-2.6}{1.1076} = -2.3475$

对给定显著性水平 $\alpha = 0.05$，可查表得 $t_{0.975}(9) = 2.2622$。

由于$|T_2| > 2.2622$，故应拒绝原假设$H_0\colon \mu = 0$，即可认为两种种子的平均单位产量有显著差异，此处检验的p值为0.0435。进一步，平均单位产量差的估计量为$\hat{u} = \overline{x} - \overline{y} = -2.6$，可见种子$Y$要比$X$的产量要高。

4.2.4 成对数据检验

我们指出成对数据T检验方法更加合理。这是因为成对数据的差D_i已消除了试验单元（如土质）之间的差别，从而用于检验的标准差$s_d = 3.5024$已排除土质差异的影响，只保留种子间的差异。

而二样本t检验中用于检验的标准差$s_w = 4.8705$还含有土质差异，从而使得标准差增大，导致因子不显著。所以成对数据场合单样本t检验所做的结论更可信些。

4.2.4 成对数据检验

- 例4.10 为了比较两种轮胎的磨损质量，设计了一个试验。把两个品牌的轮胎都分别安装在6台摩托车上并行使特定的里程。再把摩托车的前后轮胎调换，并行使相同的里程。在试验的最后，测量轮胎的磨损程度，数据如下：

	1	2	3	4	5	6
品牌A	98	61	38	117	88	109
品牌B	102	60	46	125	94	111

问：在0.05的显著性水平下，两种轮胎的磨损质量相同吗？

4.2.4 成对数据检验

- **解**：在正态性 假定下，用 $D = A - B \sim N\left(\mu, \sigma_D^2\right)$, 其中

$$\mu = \mu_1 - \mu_2, \sigma_D^2 = \sigma_1^2 + \sigma_2^2$$

要检验的假设是:

$$H_0: \mu = 0 \quad vs \quad H_1: \mu \neq 0$$

检验统计量为$T = \dfrac{\bar{D}}{S_D/\sqrt{n}}$, 拒绝域为$\{|T| \geq t_{1-\alpha/2}(n-1)\}$,

由：$\bar{d} = -4.57, n = 6,\ s_d = \left(\frac{1}{5}\sum_{i=1}^{6}(x_i - \bar{x})^2\right)^{\frac{1}{2}} = 3.56$

$$T = \frac{-4.57}{3.56/\sqrt{6}} \approx -3.096$$

查表知，$t_{0.975}(5) = 2.5706$，由于 $|T| > T_{0.975}(5)$，故拒绝原假设，认为两种轮胎有区别。

该检验的p值为：$P - value = P(|T| > 3.096) = 0.0270$

4.2.5 正态总体方差的检验

一、单个正态总体方差的χ^2检验

设$X_1,\cdots,X_n$是来自$N(\mu,\sigma^2)$的样本，对方差亦可考虑如下三个检验问题：

$$H_0:\sigma^2 \le \sigma_0^2 \quad vs \quad H_1:\sigma^2 > \sigma_0^2$$
$$H_0:\sigma^2 \ge \sigma_0^2 \quad vs \quad H_1:\sigma^2 < \sigma_0^2$$
$$H_0:\sigma^2 = \sigma_0^2 \quad vs \quad H_1:\sigma^2 \ne \sigma_0^2$$

通常假定μ未知，它们采用的检验统计量是相同的，均为$\chi^2=(n-1)S^2/\sigma_0^2$。

4.2.5 正态总体方差的检验

若取显著性水平为α，则对应三个检验问题的拒绝域依次分别为:

$$W=\{\chi^2\ge\chi^2_{1-\alpha}(n-1)\};$$
$$W=\{\chi^2\le\chi^2_{\alpha}(n-1)\};$$
$$W=\{\chi^2\le\chi^2_{\alpha/2}(n-1)\quad 或\quad \chi^2\ge\chi^2_{1-\alpha/2}(n-1)\}$$

对χ^2检验统计量代入观测值后记为χ_0^2。则对应三种假设的p值分别为:

$$p=P(\chi^2\ge\chi_0^2)$$
$$p=P(\chi^2\le\chi_0^2)$$
$$p=2\min\{P(\chi^2\ge\chi_0^2),\ P(\chi^2\le\chi_0^2)\}$$

4.2.5 正态总体方差的检验

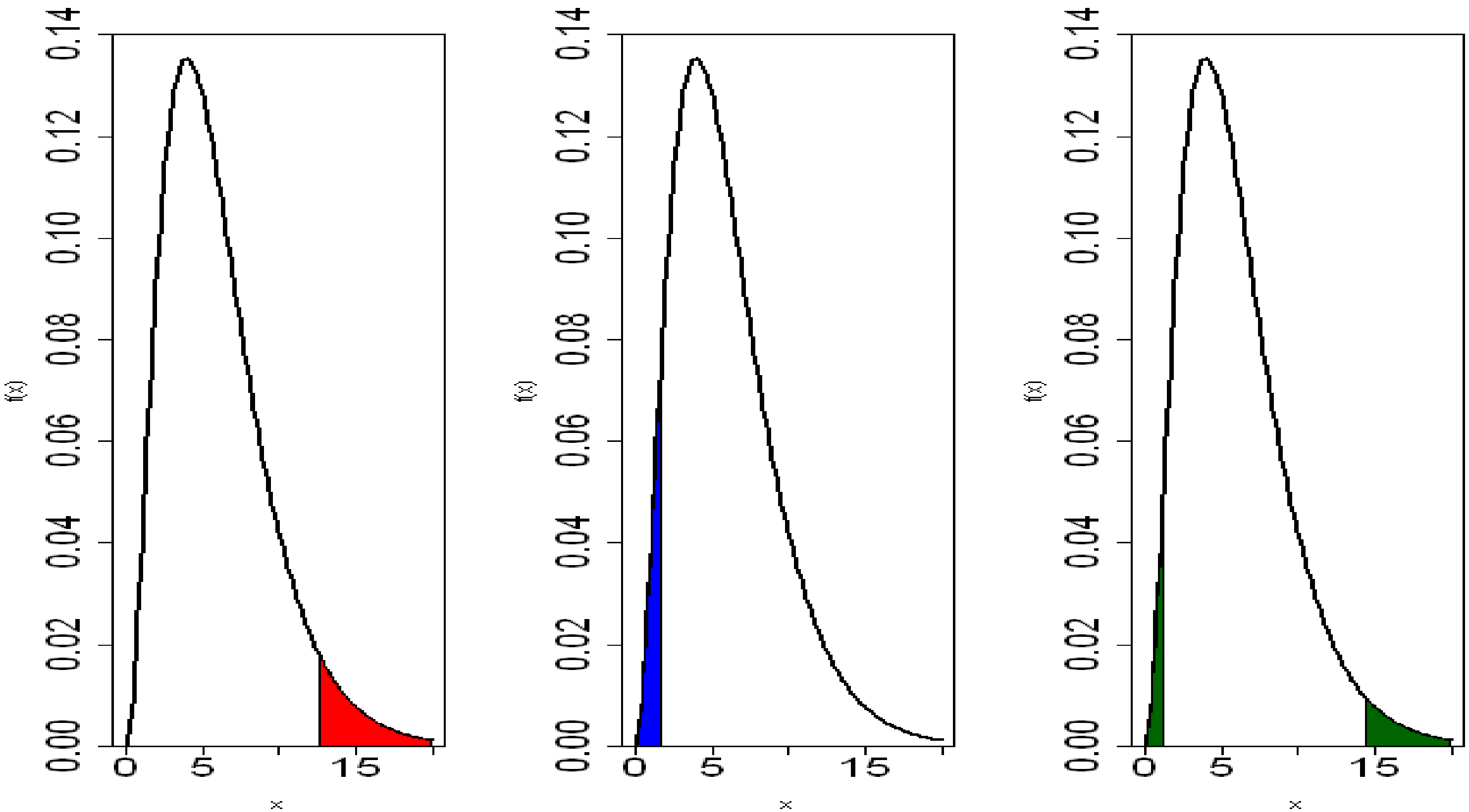


图4.4　分别对应三种假设的拒绝域示意图

(图中曲线是分布$\chi^2(6)$, $\alpha = 0.05$的密度函数曲线)

4.2.5 正态总体方差的检验

- 例4.11半导体中的某个部件对厚度要求很严格，随机抽取18个样本测量，厚度的方差为0.68。生产线如果能够把方差控制在不大于0.36即被视为合格的。假设该部件厚度服从正态分布，根据样本数据能否认为生产的部件是不合格的，以0.05为显著性水平。

4.2.5 正态总体方差的检验

❖ **解：** 原假设为 $H_0: \sigma^2 \le 0.36$,
备择假设为$H_1: \sigma^2 > 0.36$,

检验统计量为$\chi^2 = (n-1)S^2/\sigma_0^2$，其中 σ_0^2=0.36.

拒绝域为$\{\chi^2 \ge \chi^2_{1-\alpha}(n-1)\}$.

此处$n = 18$，若取 $\alpha = 0.05$，则查表知

$$\chi^2_{0.95}(17) = 27.5871$$

现计算可得：

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{17 \times 0.68}{0.36} = 32.11 > 27.5871$$

由此，在显著性水平0.05下，我们拒绝原假设，认为生产的半导体部件不符合要求。

该检验的P值为：$P - value = p\left(\chi^2 > 32.11\right) = 0.0146$

4.2.5 正态总体方差的检验

- 例4.12 一个关于正态分布总体方差的右尾检验，自由度是25，如果χ^2值为39.33，计算出P值。
- 解：该检验的P值为：

$$P - value = P\left(\chi^2 > 39.33\right) = 0.0341$$

注：查表可得到P值在0.025和0.05之间。
精确的数字可以用R语言等软件计算得到：
代码为：1-pchisq(39.33,25)

4.2.5 正态总体方差的检验

二、两个正态总体方差比的F检验

设$X_1,\cdots,X_m$是来自$N(\mu_1,\sigma_1^2)$的样本，$Y_1,\cdots,Y_n$是来自$N(\mu_2,\sigma_2^2)$ 的样本。考虑如下三个假设检验问题

$$H_0: \sigma_1^2 \le \sigma_2^2 \quad vs \quad H_1: \sigma_1^2 > \sigma_2^2$$
$$H_0: \sigma_1^2 \ge \sigma_2^2 \quad vs \quad H_1: \sigma_1^2 < \sigma_2^2$$
$$H_0: \sigma_1^2 = \sigma_2^2 \quad vs \quad H_1: \sigma_1^2 \ne \sigma_2^2$$

通常μ_1,μ_2均未知，记S_X^2，S_Y^2分别是由$X_1,\cdots,X_m$算得的σ_1^2的无偏估计和由$Y_1,\cdots,Y_n$算得的σ_2^2的无偏估计.

4.2.5 正态总体方差的检验

可建立检验统计量：$F=\frac{S_X^2}{S_Y^2}$

取显著性水平为α，对应三种假设的拒绝域分别为：

$$W=\{F\geq F_{1-\alpha}(m-1,n-1)\}$$
$$W=\{F\leq F_{\alpha}(m-1,n-1)$$
$$W=\{F\leq F_{\alpha/2}(m-1,n-1)\text{或 }F\geq F_{1-\alpha/2}(m-1,n-1)\}$$

将观测值代入F检验统计量后记为F_0，三种检验问题对应的

p值依次为

$$p=P(F\geq F_0)$$

$$p=P(F\leq F_0)$$

$$p=2\min\{P\left(F\geq F_0^2\right),\ P\left(\mathrm{F}\leq F_0^2\right)\}$$

4.2.5 正态总体方差的检验

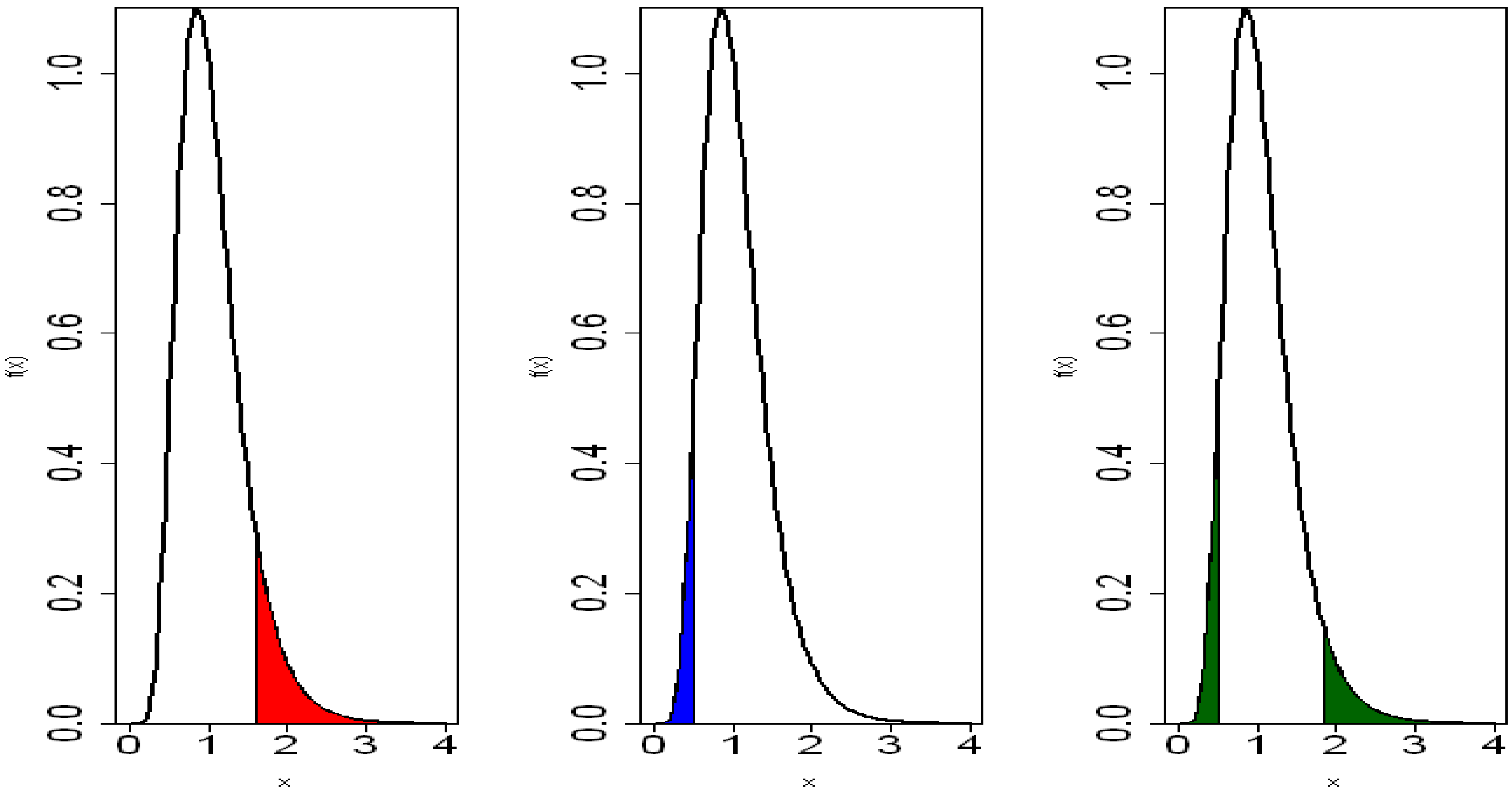


图4.5 分别对应三种假设的拒绝域示意图

(图中曲线是分布$F(20,40)$的密度函数曲线)

4.2.5 正态总体方差的检验

- 例4.13 甲、乙两台机床加工某种零件，零件的直径服从正态分布，总体方差反映了加工精度，为比较两台机床的加工精度有无差别,现从各自加工的零件中分别抽取7件产品和8 件产品，测得其直径为

X (机床甲):	16.2	16.4	15.8	15.5	16.7	15.6	15.8	
Y (机床乙):	15.9	16.0	16.4	16.1	16.5	15.8	15.7	15.0

请问根据样本观测值可以得出什么结论？（α =0. 05）

4.2.5 正态总体方差的检验

❖ **解：** $H_0: \sigma_1^2 = \sigma_2^2 \ vs \ H_1: \sigma_1^2 \neq \sigma_2^2$。

检验统计量为$F = \frac{S_X^2}{S_Y^2}$.

拒绝域为 $W = \{F \le F_{\alpha/2}(m-1, n-1)$或$F \ge F_{1-\alpha/2}(m-1, n-1)\}$.

此处 $m = 7$，$n = 8$，经计算 $s_x^2 = 0.2729$， $s_y^2 = 0.2164$,

于是$F = \frac{0.2729}{0.2164} = 1.261$，$\alpha$=0.05，查表知$F_{0.975}(6,7) = 5.12$，

$F_{0.025}(6,7) = \frac{1}{F_{0.975}(7,6)} = \frac{1}{5.70} = 0.175.$

由于$0.175 < 1.261 < 5.12$，故在0.05水平下可以认为两台机床的加工精度无显著差异。该检验的P值为：

$P = 2\min\{P(F \ge 1.261), P(F \le 1.261)\} = 2\min\{0.3967, 0.6033\} = 0.7934$。

§ 4.3 其他分布参数的假设检验

4.3.1 指数分布参数的假设检验*

设$X_1,\cdots,X_n$是来自指数分布$Exp(1/\theta)$的样本，关于θ的如下检验问题：

$$H_0:\theta\le\theta_0 \quad vs \quad H_1:\theta>\theta_0$$

拒绝域的形式是 $W=\{\bar{X}\ge c\}$，由于在$\theta=\theta_0$时，

$$\chi^2=\frac{2n\bar{X}}{\theta_0}\sim\chi^2(2n)$$

所以拒绝域为 $W=\{\chi^2\ge{\chi^2}_{1-\alpha}(2n)\}$.

4.3.1 指数分布参数的假设检验*

- **例4.14** 设我们要检验某种元件的平均寿命不小于6000小时，假定元件寿命为指数分布，现取5个元件投入试验，观测到如下5个失效时间：

  395, 4094, 119, 11572, 6133.

- **解：** 由于待检验的假设为

$$H_0: \theta \geq 6000 \quad vs \quad H_1: \theta < 6000$$

若取 α=0.05，则检验拒绝域为：

$$\{\chi^2 \leq \chi^2_{0.05}(10) = 3.94\},$$

经计算得

$$\chi^2 = \frac{10\bar{x}}{\theta_0} = \frac{10 \times 4462.6}{6000} = 7.4377$$

故不能拒绝原假设，可以认为平均寿命不低于6000小时。

该检验的P值是$P(\chi^2 \leq 7.4377) = 0.6836.$

4.3.2 比率p的检验

比例 p 可看作某事件发生的概率，对应总体可看作二点分布$b(1,p)$ 。作 n 次独立试验，设$X_1,\dots,X_n$是样本，以A记该事件发生的次数，则$A=\sum_{i=1}^{n}X_i\sim b(n,p)$。

我们可以根据 A 检验关于 p 的一些假设:

（1） $H_0: p\le p_0 \quad vs \quad H_1: p>p_0$

（2） $H_0: p\ge p_0 \quad vs \quad H_1: p<p_0$

（3） $H_0: p=p_0 \quad vs \quad H_1: p\ne p_0$

4.3.2 比率p的检验

（1） $H_0: p \le p_0 \quad vs \quad H_1: p > p_0,$

直观上看拒绝域为:$W = \{A \ge c\}$，由于A 只取整数值，故c 可限制在非负整数中。

一般情况下，对给定的α，不一定能正好取到一个正整数c 使下式成立：

$$P(A \ge c; p_0) = \sum_{i=c}^{n} \binom{n}{i} {p_0}^i (1-p_0)^{n-i} = \alpha,$$

这是在对离散总体作假设检验中普遍会遇到的问题。

4.3.2 比率p的检验

一般较常见的是找一个c_0，使得

$$\sum_{i=c_0+1}^{n}\binom{n}{i}{p_o}^{i}(1-p_0)^{n-i}<\alpha<\sum_{i=c_0}^{n}\binom{n}{i}{p_o}^{i}(1-p_0)^{n-i}$$

得到拒绝域为: $W=\{A\geq c_0+1\}$.

事实上，在离散场合使用p值作检验较为简便。这时可以不用找c_0,只需根据观测值$A=x_0$计算检验的p值:

$$p=P_{p_0}\{A\geq x_0\},$$

并将之与显著性水平比较大小即可。

4.3.2 比率p的检验

（2） $H_0: p \geq p_0 \quad vs \quad H_1: p < p_0$

检验的拒绝域为: $W = \{A \leq c\}$，c 为满足下式的最大正整数:

$$\sum_{i=0}^{c} \binom{n}{i} {p_0}^i (1-p_0)^{n-i} \leq \alpha$$

p值$= P_{p_0}(A \leq x_0)$。

4.3.2 比率p的检验

（3） $H_0: p = p_0 \quad vs \quad H_1: p \neq p_0$

检验的拒绝域为:$W = \{A \le c_1$或$A \ge c_2\}$， 其中，c_1为满足下式的最大正整数:

$$\sum_{i=0}^{c_1} \binom{n}{i} {p_0}^i (1-p_0)^{n-i} \le \frac{\alpha}{2}$$

c_2为满足下式的最小正整数:

$$\sum_{i=c_2}^{n} \binom{n}{i} {p_0}^i (1-p_0)^{n-i} \le \frac{\alpha}{2}$$

p值$= 2min\{P_{p_0}(A \le x_0), P_{p_0}(A \ge x_0)\}$。

4.3.2 比率p的检验

- 如在产品质量检查中检察员想要知道不合格率p是否高于10%，抽取了40件样品，发现有8件不合格，那么根据抽检结果检查员能够得出什么结论？

- 总体是服从$b(1,p)$。设X为n次抽样中不合格的产品数量，则有$X \sim b(n,p)$，其中 $n=40$，X 的观测值 $x_0=8$。

要检验的假设为：

$$H_0\!: p \le 10\% \quad vs \quad H_1\!: p > 10\%$$

4.3.2 比率p的检验

$$p-value = P_{p_0}(X \geq x_0) = P_{p_0}(X \geq 8)$$

$$= \sum_{i=8}^{n} \binom{n}{i} 0.1^i (1-0.1)^{n-i}$$

$$= 1 - 0.9^{40} - \binom{40}{1} 0.1 \times 0.9^{39} - \cdots - \binom{40}{7} 0.1^7 \times 0.9^{33}$$

$$= 0.0419.$$

若取α=0.05，由于$p < \alpha$，则应拒绝原假设。

4.3.2 比率p的检验

- 例4.15 某厂生产的产品优质品率一直保持在40%，近期对该厂生产的该类产品抽检 20件，其中优质品7件，在$\alpha = 0.05$下能否认为优质品率仍保持在40%?

- 解：以p 表示优质品率，X表示20件产品中的优质品件数，则$X \sim b(20, p)$，待检验的假设为

$$H_0: p = 0.4 \quad vs \quad H_1: p \neq 0.4$$

拒绝域为

$$W = \{X \leq c_1 \text{或} X \geq c_2\}$$

4.3.2 比率p的检验

下面求c_1与c_2:由于

$P_{p_0}(X \le 3) = 0.0160 < 0.025 < P_{p_0}(X \le 4) = 0.0510,$

故取 c_1=3，又因为

$P_{p_0}(X \ge 11) = 0.0565 > 0.025 > P_{p_0}(X \ge 12) = 0.0210,$

从而c_2=12，因此，拒绝域为$W = \{X \le 3$或$X \ge 12\}$.

由于观测值没有落入拒绝域，故不能拒绝原假设。

附带指出，该拒绝域的显著性水平实际上不是0.05，而是0.0160+0.021=0.0370 。

4.3.2 比率p的检验

- 另解：$X \sim b(20, p)$，待检验的假设为

$$H_0: p = 0.4 \quad vs \quad H_1: p \neq 0.4$$

拒绝域为$W = \{X \le c_1$或$X \ge c_2\}$.

p值$= 2\min\{P_{p_0}(X \le x_0), P_{p_0}(X \ge x_0)\}$, x_0=7,

$P_{p_0}(X \ge 7) = 1 - P(X \le 6) = 0.7500$

$P_{p_0}(X \le 7)$ =0.4159

从而 p值 = 2× 0.4159=0.8318.

由于p值 $> \alpha = 0.05$，故不能拒绝原假设。

4.3.3 大样本检验

- 在二点分布参数 p 的检验问题中，临界值的确定比较繁琐，使用不太方便。如果样本量较大，我们可用近似的检验方法——大样本检验。
- 大样本检验一般思路如下：设$X_1,\ldots,X_n$是来自某总体的样本，又设该总体均值为θ，方差为θ的函数，记为$\sigma^2(\theta)$。譬如，对二点分布$b(1,\theta)$，其方差 $\theta(1-\theta)$是均值θ的函数，则在样本容量 n 充分大时，

$$\bar{X}\dot{\sim}N(\theta,\sigma^2(\theta)/n)$$

故在$\theta=\theta_0$时，可采用如下检验统计量

$$U=\frac{\sqrt{n}(\bar{X}-\theta_0)}{\sqrt{\sigma^2(\hat{\theta})}}\dot{\sim}N(0,1)$$

由此近似地确定拒绝域。其中，$\hat{\theta}$ 是θ的MLE.

4.3.3 大样本检验

- 例4.16 某厂产品的不合格品率为 10%，在 一次例行检查中，随机抽取80件，发现有11件不合格品，在α =0.05下能否认为不合格品率仍为10%？
- 解：这是关于不合格品率的检验，假设为:

$$H_0: \theta = 0.1 \quad vs \quad H_1: \theta \neq 0.1$$

因为n =80 比较大，可采用大样本检验方法。

检验统计量为 $U = \dfrac{\sqrt{n}(\bar{X}-0.1)}{\sqrt{\hat{\theta}(1-\hat{\theta})}}$,

拒绝域为$W = \{|U| \geq u_{1-\alpha/2}\}$.

4.3.3 大样本检验

$\hat{\theta} = \frac{11}{80}$, 检验统计量为

$$U = \frac{\sqrt{80}\left(\frac{11}{80} - 0.1\right)}{\sqrt{0.1186}} = 0.9739$$

若取α=0.05，则$u_{0.975}$=1.96, 故拒绝域为

$$W = \{|U| \geq 1.96\},$$

故不能拒绝原假设。

该检验P值为：

$p - value = P(|U| > 0.9739) = 0.3301.$

4.3.3 大样本检验

- 例 4.17 某建筑公司宣称其麾下建筑工地平均每天发生事故数不超过 0.6 起，现记录了该公司麾下建筑工地 200天的安全生产情况，事故数记录如下：

一天发生的事故数	0	1	2	3	4	5	≥ 6	合计
天数	102	59	30	8	0	1	0	200

试检验该建筑公司的宣称是否成立(取α=0.05)。

4.3.3 大样本检验

- **解：** 以 X 记建筑工地一天发生的事故数，可认为 $X \sim P(\lambda)$，要检验的假设是：

$$H_0: \lambda \le 0.6 \quad vs \quad H_1: \lambda > 0.6$$

由于n=200很大，可以采用大样本检验。

泊松分布的均值和方差都是λ,

检验统计量为$U = \frac{\sqrt{n}(\bar{X}-0.6)}{\sqrt{\hat{\lambda}}}$, 其中$\hat{\lambda} = \bar{X}$.

拒绝域为$W = \{U \ge u_{1-\alpha}\}$.

4.3.3 大样本检验

这里$\bar{x}=0.74$，检验统计量观测值为

$$u_0=\frac{\sqrt{n}(\bar{x}-\lambda)}{\sqrt{\hat{\lambda}}}=\frac{\sqrt{200}(0.74-0.6)}{\sqrt{0.74}}=2.302.$$

若取α=0.05，则 $u_{0.95}$=1.645，拒绝域为

$$W=\{U\geq 1.645\}$$

如今 u_0=2.302 已落入拒绝域，故拒绝原假设，认为该建筑公司的宣称明显不成立。

该检验P值为：$p-value=P(U>2.302)=0.0102.$

4.3.3 大样本检验

- 例4.18 一个备择假设为$p > p_0$的检验，大样本检验统计量的值为2.18，则该检验的p值是多少？

- 解：该检验p值为：

$$p - value = P_{p_0}(U > 2.18) = 0.0146$$

4.3.3 大样本检验

大样本检验是近似的:

- 近似的含义是指检验的实际显著性水平与原先设定的显著性水平有差距，这是由于诸如(4.3.6)中U 的分布与$N(0,1)$有距离。如果n 很大，则这种差异就很小。

- 实用中我们一般并不清楚对一定的n, U 的分布与$N(0,1)$的差异有多大，因而也就不能确定检验的实际水平与设定水平究竟差多少。在区间估计中也有类似问题。

- 因此，大样本方法是一个“不得已而为之”的方法。只要有基于精确分布的方法一般总是首先要加以考虑的。

§4.4 似然比检验与分布拟合检验

4.4.1 似然比检验的思想*

- 我们在前几节讲述的内容均是关于费希尔提出的显著性检验，类似于在估计中存在着多种估计一样，在假设检验中，也有多种检验方法，如奈曼（Neyman）和皮尔逊（E.Pearson）于1928年提出的似然比检验，它是一种应用较广的检验方法。

4.4.1 似然比检验的思想*

- 定义4.4 设$X_1,\cdots,X_n$为来自密度函数为$f(x;\theta),\theta\in\Theta$的样本，考虑如下检验问题

$$H_0:\theta\in\Theta_0\ \ vs\ \ H_1:\theta\in\Theta_1=\Theta-\Theta_0$$

令

$$\Lambda(x_1,\cdots,x_n)=\frac{\sup\limits_{\theta\in\Theta} f(x_1,\cdots,x_n;\theta)}{\sup\limits_{\theta\in\Theta_0} f(x_1,\cdots,x_n;\theta)}$$

则我们称$\Lambda(x_1,\cdots,x_n)$为上述假设的似然比（likelihood ratio），有时也称之为广义似然比。

4.4.1 似然比检验的思想*

上式的$\Lambda(x_1,\cdots,x_n)$也可以写成如下形式

$$\Lambda(x_1,\cdots,x_n)=\frac{f(x_1,\cdots,x_n;\hat{\theta})}{f(x_1,\cdots,x_n;\hat{\theta}_0)}$$

其中$\hat{\theta}$ 表示在全参数空间Θ上θ的最大似然估计，$\hat{\theta}_0$表示在子参数空间Θ_0上θ的最大似然估计。也就是说，$\Lambda(x_1,\cdots,x_n)$的分子表示没有假设时的似然函数最大值，分母表示在原假设成立条件下的似然函数最大值，不难看出，如果$\Lambda(x_1,\cdots,x_n)$的值很大，则说明$\theta\in\Theta_0$的可能性要比$\theta\in\Theta_1$的可能性小，于是我们有理由认为H_0不成立。这样，我们有如下的似然比检验。

4.4.1 似然比检验的思想*

❖ 定义4.5 当采用似然比统计量

$$\Lambda(x_1,\cdots,x_n)=\frac{f(x_1,\cdots,x_n;\hat{\theta})}{f(x_1,\cdots,x_n;\hat{\theta}_0)}$$

作为定义4.4 中检验问题的检验统计量，且取其拒绝域为 $W=\{\Lambda(x_1,\cdots,x_n)\geq c\}$，其中临界值$c$满足

$$P_\theta(\Lambda(x_1,\cdots,x_n)\geq c)\leq\alpha,\qquad \forall\theta\in\Theta_0$$

则称此检验为显著性水平α的似然比检验（likelihood ratio test），简记为LRT.

4.4.1 似然比检验的思想*

- 例4.19 设$X_1,\cdots,X_n$是来自正态总体$N(\mu,\sigma^2)$的样本，μ,σ^2均未知。试求检验问题

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

的显著性水平为α的似然比检验。

- 解：记$\theta=(\mu,\sigma^2)$，样本联合密度函数为

$$f(x_1,\cdots,x_n;\theta)=(2\pi\sigma^2)^{-\frac{n}{2}}exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\},$$

两个参数空间分别为

$$\Theta_0=\{(\mu_0,\sigma^2)|\sigma^2>0\},\Theta=\{(\mu,\sigma^2)|\mu\in R,\sigma^2>0\}$$

4.4.1 似然比检验的思想*

利用微分法，我们容易求得，

在Θ上， $\hat{\mu}=\bar{X}, \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$分别为$\mu$与$\sigma^2$的MLE，

在Θ_0上， $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2$是$\sigma^2$的MLE，

代回各自似然函数后，可得

$$\sup_{\theta\in\Theta_0} f(x_1,\cdots,x_n;\theta)=\left[2\pi\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2\right]^{-n/2} e^{-n/2}$$

$$\sup_{\theta\in\Theta} f(x_1,\cdots,x_n;\theta)=\left[2\pi\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{-n/2} e^{-n/2}$$

4.4.1 似然比检验的思想*

于是，其似然比统计量为

$$\Lambda(x_1,\cdots,x_n)=\frac{\sup\limits_{\theta\in\Theta} f(x_1,\cdots,x_n;\theta)}{\sup\limits_{\theta\in\Theta_0} f(x_1,\cdots,x_n;\theta)}=\left(\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^{n/2}$$

$$=\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^{n/2}=\left(1+\frac{T^2}{n-1}\right)^{n/2}$$

其中$T=\frac{\sqrt{n}(\bar{X}-\mu_0)}{S}$就是4.2节中的$T$检验统计量。

4.4.1 似然比检验的思想*

从上式可知，此时的似然比统计量Λ是传统的T统计量平方的严增函数，于是，两个检验统计量的拒绝域有如下等价关系：

$$\{\Lambda(x_1,\cdots,x_n)\geq c\}\leftrightarrow\{|T|\geq d\}$$

且由T的分位数可定出Λ的分位数。

又因为当原假设成立时，$T\sim t(n-1)$，若我们取$d=t_{1-\frac{\alpha}{2}}(n-1)$则用$c=\left[1+\frac{d^2}{n-1}\right]^{n/2}$ 就可控制用 Λ犯第一类错误的概率不超过α，由此可见，此时的似然比检验与我们前面讲过的双侧T检验完全等价。

4.4.2 分类数据的χ^2拟合优度检验

设总体X 可以分成r 类，记为$A_1, \cdots, A_r$，现对该总体作了n次观测，r 个类出现的频数分别为$n_1, \ldots, n_r$且 $\sum_{i=1}^{r} n_i = n$

检验如下假设:

$$H_0: \ P(A_i) = p_{i0}, \quad i = 1, 2, \cdots, r.$$

$$H_1: \text{存在 } i, \ P(A_i) \neq p_{i0}.$$

其中诸$p_{i0} \geq 0$且$\sum_{i=1}^{r} p_{i0} = 1$.

4.4.2 分类数据的χ^2拟合优度检验

(1) 诸 p_{i0} 均已知（多项分布的大样本检验）

如果H_0 成立，则对每一类A_i，其频率n_i/n与概率p_{i0}应较接近。即观测频数n_i与理论频数np_{i0}应相差不大。

据此，英国统计学家$K.Pearson$提出如下检验统计量:

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}}$$

并证明在H_0 成立时对充分大的n，上式给出的检验统计量近似服从自由度为$r-1$的χ^2分布。

拒绝域为: $W=\{\chi^2\geq\chi^2_{1-\alpha}(r-1)\}$

4.4.2 分类数据的χ^2拟合优度检验

- 例4.20 在19世纪，孟德尔按颜色和形状把豌豆分为四类：黄圆、绿圆、黄皱、绿皱。孟德尔根据遗传学原理判断这四类的比例应为9:3:3:1。为做验证，孟德尔在一次豌豆实验中收获了$n = 556$个豌豆，其中这四类豌豆的个数分别为315,108,101,32。该数据是否与孟德尔提出的比例吻合？

4.4.2 分类数据的χ^2拟合优度检验

- **解：** 对上述数据我们可以做如下的χ^2拟合优度检验。注意到此时
$r=4, n=556, n_1=315, n_2=108, n_3=101, n_4=32,$
待检验的假设为

$$H_0: p_{10}=\frac{9}{16}, p_{20}=\frac{3}{16}, p_{30}=\frac{3}{16}, p_{40}=\frac{1}{16}.$$

检验统计量为 $\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}},$
拒绝域为$\{\chi^2 \geq \chi^2_{1-\alpha}(r-1)\}$。
计算出np_{i0}故有

$$\chi^2=\frac{(315-312.75)^2}{312.75}+\frac{(108-104.25)^2}{104.25}+\frac{(101-104.25)^2}{104.25}+\frac{(32-34.75)^2}{34.75}=0.47$$

4.4.2 分类数据的χ^2拟合优度检验

若取α=0.05，则$\chi^2_{0.95}(3)=7.8147>0.47$,故没有理由拒绝$H_0$，即认为孟德尔的结论是可以接受的。

该检验的近似p值也是可以计算的，为

$$p=P\left(\chi^2\geq 0.47\right)=0.9254$$

其中χ^2表示服从 $\chi^2(3)$的随机变量。因此，我们不能拒绝原假设。

4.4.3 分布的χ^2拟合优度检验

（2）诸 p_{i0} 不完全已知（其它分布的大样本检验）

H_0：总体X服从指定的分布

H_1：总体X不服从该分布

指定的分布可以是离散分布，也可以是连续分布。

离散分布情形：总体X的取值是有限的或可列的，把这些取值分为若干类。统计样本观测值落入每一类的频数，得到实际频数。

连续分布情形：总体X的取值是连续的，同样对X的取值分类，类似可得到样本观测值落入每一类的实际频数。

假设中指定的分布可以是完全已知的分布，也可以是包含有有限个未知参数的分布。

4.4.3 分布的χ^2拟合优度检验

若假设中指定的分布包含有有限个未知参数，也就是说：诸 $p_{i0}, i=1,\cdots,r$ 由$k(k<r)$个未知参数$\theta_1,\ldots,\theta_k$确定，即 $p_{i0}=p_{i0}(\theta_1,\cdots,\theta_k)$, $\quad i=1,\cdots,r.$

首先给出$\theta_1,\cdots,\theta_k$的极大似然估计$\hat{\theta}_1,\cdots,\hat{\theta}_k$,然后给出诸 $p_{i0}, i=1,\cdots,r$的极大似然估计$\hat{p}_{i0}=p_{i0}\left(\hat{\theta}_1,\ldots,\hat{\theta}_k\right)$.

Fisher证明了在H_0成立时$\chi^2=\sum_{i=1}^{r}\frac{(n_i-n\hat{p}_{i0})^2}{n\hat{p}_{i0}}$近似服从自由度为$r-k-1$的$\chi^2$分布，于是检验拒绝域为:

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

4.4.3 分布的χ^2拟合优度检验

❖ **例4.21 卢瑟福在2608个等时间间隔内观测一枚放射性物质放射的粒子数X，表4.5是观测结果的汇总，其中n_k表示2608次观测中放射粒子数为k的次数。**

表4.5　卢瑟福实验数据

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n_k	57	203	383	525	532	408	273	139	45	27	10	4	2	0	0

试利用该组数据检验该放射物质在单位时间内放射出的粒子数是否服从泊松分布。

4.4.3 分布的χ^2拟合优度检验

- **解：** 本例中，要检验总体是否服从泊松分布。

$$H_0: \text{质点数服从泊松分布 } P(\lambda)$$

$$H_1: \text{质点数不服从泊松分布} P(\lambda)$$

观测到 0, 1, …, 14共15个不同取值。

首先估计未知参数λ，采用极大似然估计，

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\sum_{k=0}^{14} k n_k}{\sum_{k=0}^{14} n_k}$$

$$= \frac{1}{2608}(1 \times 203 + 2 \times 383 + \ldots + 11 \times 4 + 12 \times 2) = 3.87$$

将$\hat{\lambda}$代入可以估计出诸$\hat{p}_k$, $\hat{p}_k = \frac{\hat{\lambda}^k}{k!}e^{-\hat{\lambda}}, k = 0, 1, 2, \cdots$

4.4.3 分布的χ^2拟合优度检验

列表如下：

i	n_i	$\hat{p}_i$	$n\hat{p}_i$	$(n_i - n\hat{p}_i)^2/n\hat{p}_i$
0	57	0.0209	54.5	0.1147
1	203	0.0807	210.5	0.2672
2	383	0.1562	407.4	1.4614
3	525	0.2015	525.5	0.0005
4	532	0.1950	508.6	1.0766
5	408	0.1509	393.5	0.5343
6	273	0.0973	253.8	1.4525
7	139	0.0538	140.3	0.0120
8	45	0.0260	67.8	7.6673
9	27	0.0112	29.2	0.1658
10	10	0.0043	11.2	0.1258
≥11	6	0.0022	5.7	0.0158
合计	2608	1.0000	2068	$\chi^2 = 12.8967$

4.4.3 分布的χ^2拟合优度检验

为了满足每一类出现的样本观测次数不小于5，我们把$k \geq 11$作为一类，记为第12类，则检验统计量为

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

拒绝域为$\{\chi^2 \geq \chi^2_{1-\alpha}(r-1-1)\}$，其中$r = 12$.

计算结果列在上表中, 可以看到检验统计量的值为

$$\chi^2 = \sum_{i=1}^{12} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = 12.8967.$$

4.4.3 分布的χ^2拟合优度检验

若取α=0.05，则$\chi^2_{1-\alpha}(r-k-1)=\chi^2_{0.95}(10)=18.307$.

χ^2=12.8967<18.307，故接受原假设，可以认为该放射性物质在单位时间里放出的质点数服从泊松分布。

使用统计软件可以计算出此处检验的p 值是0.2295。

4.4.3 分布的χ^2拟合优度检验

- 例4.22 某工厂生产一种滚珠，现随机抽取50件产品，测得其直径为（单位：mm）

15.0	15.8	15.2	15.1	15.9	14.7	14.8	15.5	15.6	15.3
15.0	15.6	15.7	15.8	14.5	15.1	15.3	14.9	14.9	15.2
15.9	15.0	15.3	15.6	15.1	14.9	14.2	14.6	15.8	15.2
15.2	15.0	14.9	14.8	15.1	15.5	15.5	15.1	15.1	15.0
15.3	14.7	14.5	15.5	15.0	14.7	14.6	14.2	14.2	14.5

问滚珠直径是否服从正态分布？

4.4.3 分布的χ^2拟合优度检验

- **解：** H_0：滚珠直径 X 服从正态分布$N(\mu,\sigma^2)$

  H_1：滚珠直径 X 不服从正态分布$N(\mu,\sigma^2)$

μ,σ^2的MLE为 $\hat{\mu}=\bar{x}=15.1$， $\hat{\sigma}^2=s_n^2=0.4325^2$.

对观测数据分组，

数据最大值是15.9，最小值是14.2.

组数设为5，组距设为0.4，

若分点取为

$a_0=-\infty, a_1=14.55,$

$a_2=14.95,\ a_3=15.35,$

$a_4=15.75, a_5=+\infty.$

则可作出样本观测值频数分布表。

组号	分组	频数
1	$(-\infty, 14.55]$	6
2	$(14.55, 14.95]$	11
3	$(14.95, 15.35]$	20
4	$(15.35, 15.75]$	8
5	$(15.75, +\infty]$	5
总和		50

4.4.3 分布的χ^2拟合优度检验

组号 i	分组	实际频数 n_i	事件	估计概率 $\hat{p}_i$	理论频数 $n\hat{p}_i$	$\frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
1	$(-\infty, 14.55]$	6	$X \le 14.55$	0.10174	5.0872	0.1638
2	(14.55, 14.95]	11	$X \le 14.95$ ≥ 14.55	0.26262	13.1309	0.3458
3	(14.95, 15.35]	20	$X \le 14.95$ ≥ 15.35	0.35402	17.7008	0.2986
4	(15.35, 15.75]	8	$X \le 15.35$ ≥ 15.75	0.21519	10.7593	0.7077
5	(15.75, +∞]	5	$X \ge 15.75$	0.06643	3.3217	0.8480
总和		50		1	50	2.3639

注：表中$\hat{p}_i = \Phi\left(\frac{a_i-\hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{a_{i-1}-\hat{\mu}}{\hat{\sigma}}\right), i = 1,2,3,4,5.$

$\hat{\mu} = 15.1$，$\hat{\sigma} = 0.4325.$

4.4.3 分布的χ^2拟合优度检验

从表中我们可以看到检验统计量的值为

$$\chi^2=\sum_{i=1}^{5}\frac{(n_i-n\widehat{p_i})^2}{n\widehat{p_i}}=2.3639.$$

若取α=0.05，则$\chi^2_{1-\alpha}(r-k-1)=\chi^2_{0.95}(5-2-1)=5.9915.$

$2.3639<5.9915$，故不能拒绝原假设，在这种分组数据情形下可以认为滚珠直径服从正态分布。

分组不同，则组数和对应的分点也不同，要检验的分类概率也会随之变化，因此若对检验结论有怀疑时，可用不同的分组进行尝试。显著性检验看重拒绝，若有一种分组检验结论为拒绝，则认为关于该分布的假设不成立。

4.4.4 列联表的独立性检验

- 列联表是将观测数据按两个或更多属性 (定性变量) 分类时所列出的频数表。
- 例如，对随机抽取的1000人按性别（男或女）及色觉 (正常或色盲) 两个属性分类, 得到如下二维列联表, 又称2×2表或四格表。

性别	视觉	
	正常	色盲
男	535	65
女	382	18

4.4.4 列联表的独立性检验

- 一般,若总体中的个体可按两个属性$\boldsymbol{A}$与$\boldsymbol{B}$分类，$\boldsymbol{A}$有$\boldsymbol{r}$ 个类$\boldsymbol{A_1,\cdots,A_r}$，$\boldsymbol{B}$ 有$\boldsymbol{c}$个类$\boldsymbol{B_1,\cdots,B_c}$，从总体中抽取大小为$\boldsymbol{n}$的样本，设其中有$\boldsymbol{n_{ij}}$个个体既属于$\boldsymbol{A_i}$类又属于$\boldsymbol{B_j}$类，$\boldsymbol{n_{ij}}$ 称为频数，将$\boldsymbol{r\times c}$个$\boldsymbol{n_{ij}}$排列为一个$\boldsymbol{r}$行$\boldsymbol{c}$列的二维列联表，简称$\boldsymbol{r\times c}$列联表（表4.6）。

4.4.4 列联表的独立性检验

表4.6　$r \times c$列联表

$A\backslash B$	1	$\cdots$	j	$\cdots$	c	和
1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1c}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{ic}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	n_{r1}	$\cdots$	n_{rj}	$\cdots$	n_{rc}	$n_{r\cdot}$
和	$n_{\cdot 1}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot c}$	n

4.4.4 列联表的独立性检验

- 列联表分析的基本问题是：考察各属性之间有无关联，即判别两属性是否独立。
- 如在前例中，问题是：一个人是否色盲与其性别是否有关？在$r \times c$表中，若以$p_{i\cdot}, p_{\cdot j}$和p_{ij}分别表示总体中的个体仅属于A_i，仅属于B_j和同时属于A_i与B_j的概率，可得一个二维离散分布表（表4.7）

4.4.4 列联表的独立性检验

表4.7 二维离散分布表

$A\backslash B$	1	$\cdots$	j	$\cdots$	c	行和
1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1c}	$p_{1\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	p_{ic}	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	p_{r1}	$\cdots$	p_{rj}	$\cdots$	p_{rc}	$p_{r\cdot}$
列和	$p_{\cdot 1}$	$\cdots$	$p_{\cdot j}$	$\cdots$	$p_{\cdot c}$	p

4.4.4 列联表的独立性检验

- 则"A、B两属性独立"的假设可以表述为：

$$H_0: p_{ij}=p_{i\cdot}p_{\cdot j}, \qquad i=1,\cdots,r,\ j=1,\cdots,c$$

这就变为上一小节中诸p_{ij}不完全已知时的分布拟合检验。这里诸p_{ij}共有rc个参数，在原假设H_0成立时，这rc个参数p_{ij}由$r+c$个参数$p_{1\cdot},\ldots,p_{r\cdot}$和$p_{\cdot 1},\ldots,p_{\cdot c}$决定。在这$r+c$个参数中存在两个约束条件：

$$\sum_{i=1}^{r} p_{i\cdot}=1, \qquad \sum_{j=1}^{c} p_{\cdot j}=1$$

所以，此时p_{ij}实际上由r+c−2个独立参数所确定。

4.4.4 列联表的独立性检验

据此，检验统计量为

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij}-n\hat{p}_{ij}\right)^2}{n\hat{p}_{ij}}$$

在$\boldsymbol{H_0}$成立时，上式服从自由度为$\boldsymbol{rc-(r+c-2)-1=(r-1)(c-1)}$的$\boldsymbol{\chi^2}$分布。其中诸$\boldsymbol{\hat{p}_{ij}}$是在$\boldsymbol{H_0}$成立下得到的$\boldsymbol{p_{ij}}$的极大似然估计，其表达式为：

$$\boldsymbol{\hat{p}_{ij}=\hat{p}_{i\cdot}\,\hat{p}_{\cdot j}=\frac{n_{i\cdot}}{n}\cdot\frac{n_{\cdot j}}{n}}$$

对给定的显著性水平$\boldsymbol{\alpha}$，检验的拒绝域为:

$$\boldsymbol{W=\{\chi^2\geq\chi^2_{1-\alpha}((r-1)(c-1))\}.}$$

4.4.4 列联表的独立性检验

- 例4.23 为研究儿童智力发展与营养的关系，某研究机构调查了1436名儿童，得到如表4.8 的数据，试在显著性水平0.05下判断智力发展与营养有无关系。

表4.8 儿童智力与营养的调查数据

	智商				合计
	<80	80~90	90~99	>100	
营养良好	367	342	266	329	1304
营养不良	56	40	20	16	132
合计	423	382	286	345	1436

4.4.4 列联表的独立性检验

- **解**：用A表示营养状况，它有两个水平:A_1表示营养良好，A_2表示营养不良；B表示儿童智商,它有四个水平,B_1,B_2,B_3,B_4分别表示表中四种情况。沿用前面的记号，首先建立假设:

  H_0：营养状况与智商无关联，即A与B独立

  H_1：营养状况与智商有关联，即A与B不独立

统计表示如下：

$$H_0: p_{ij} = p_{i.}p_{.j},\quad i = 1,2,\quad j = 1,2,3,4.$$

$$H_1: 存在i，j，p_{ij} = p_{i.}p_{.j}不成立$$

4.4.4 列联表的独立性检验

检验统计量可取为

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij}-n\hat{p}_{ij}\right)^2}{n\hat{p}_{ij}},$$

其中$\hat{p}_{ij}=\hat{p}_{i\cdot}\,\hat{p}_{\cdot j}=\frac{n_{i\cdot}}{n}\cdot\frac{n_{\cdot j}}{n}$.

对给定的显著性水平α，检验的拒绝域为:

$$W=\left\{\chi^2\geq\chi^2_{1-\alpha}\left((r-1)(c-1)\right)\right\},\qquad r=2,c=4.$$

4.4.4 列联表的独立性检验

在原假设H_0成立下，我们可以计算诸参数的极大似然估计值:

$$\hat{p}_{1\cdot} = 1304/1436 = 0.9081, \quad \hat{p}_{2\cdot} = 132/1436 = 0.0919,$$

$$\hat{p}_{\cdot 1} = 423/1436 = 0.2946, \quad \hat{p}_{\cdot 2} = 382/1436 = 0.2660,$$

$$\hat{p}_{\cdot 3} = 286/1436 = 0.1992, \quad \hat{p}_{\cdot 4} = 345/1436 = 0.2403,$$

进而可给出诸$\boldsymbol{n\hat{p}_{ij} = n\hat{p}_{i\cdot}\hat{p}_{\cdot j}}$，如

$$\boldsymbol{n\hat{p}_{11} = 1436 \times 0.9081 \times 0.2496 = 384.1677}$$

其它结果见表4.9。

表4.9 诸$\boldsymbol{n\hat{p}_{ij}}$的计算结果

	<80	80~90	90~99	>100	$\hat{p}_{i\cdot}$
营养良好	384.1677	346.8724	259.7631	313.3588	0.9081
营养不良	38.8779	35.1036	26.2881	31.7120	0.0919
$p_{\cdot j}$	0.2946	0.2660	0.1992	0.2403	

4.4.4 列联表的独立性检验

由表4.8 和表4.9 可以计算检验统计量的值

$$\chi^2=\frac{(367-384.1677)^2}{384.1677}+\frac{(342-346.8724)^2}{346.8724}+\cdots$$

$$+\frac{(16-31.7120)^2}{31.7120}=19.2785$$

此处$r=2$，$c=4$，$(r-1)(c-1)=3$,若取$\alpha=0.05$，查表有 $\chi^2_{0.95}(3)=7.815$，由于19.2785>7.815，故拒绝原假设，认为营养状况对智商有影响。本例中检验的p 值为0.0002。

4.4.4 列联表的独立性检验

- 例4.24 一项关于宗教信仰与对待生育控制态度之间关系的研究正在进行。 当地的三大主流宗教被作为样本。使用如下数据判断宗教信仰与对待生育控制的态度是否独立。使用0.01作为显著性水平。

表4.10　宗教数据

		宗教信仰		
		A	B	C
态度	赞成生育控制	123	64	43
	反对生育控制	77	86	57

4.4.4 列联表的独立性检验

- **解**：记宗教信仰为A，有三个水平:A_1，A_2，A_3；记态度为B，有两个水平：B_1，B_2。

  沿用前面的记号，首先建立假设:

  H_0：宗教信仰与对待生育控制的态度是独立的

  H_1：宗教信仰与对待生育控制的态度不独立

统计表示如下：

$$H_0: p_{ij} = p_{i.}p_{.j}, \qquad i = 1,2, \quad j = 1,2,3.$$

$$H_1: \text{存在} i，j，p_{ij} = p_{i.}p_{.j} \text{不成立}$$

4.4.4 列联表的独立性检验

检验统计量可取为

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij}-n\hat{p}_{ij}\right)^2}{n\hat{p}_{ij}},$$

其中$\hat{p}_{ij}=\hat{p}_{i\cdot}\,\hat{p}_{\cdot j}=\frac{n_{i\cdot}}{n}\cdot\frac{n_{\cdot j}}{n}.$

对给定的显著性水平α，检验的拒绝域为:

$$W=\left\{\chi^2\geq\chi^2_{1-\alpha}\left((r-1)(c-1)\right)\right\},\qquad r=2,c=3.$$

4.4.4 列联表的独立性检验

在原假设H_0成立下，我们可以计算诸参数的极大似然估计值:

$$\hat{p}_{\cdot 1} = 200/450 = 0.4444, \quad \hat{p}_{\cdot 2} = 150/450 = 0.3333,$$
$$\hat{p}_{\cdot 3} = 100/450 = 0.2222,$$
$$\hat{p}_{1\cdot} = 230/450 = 0.5111, \hat{p}_{2\cdot} = 220/450 = 0.4889$$

进而可给出诸$\boldsymbol{n\hat{p}_{ij} = n\hat{p}_{i\cdot}\hat{p}_{\cdot j}}$，如

$$\boldsymbol{n\hat{p}_{11} = 450 \times 0.4444 \times 0.5111 = 102.2222}$$

其它结果见表4.11。

4.4.4 列联表的独立性检验

表4.11　诸$n\hat{p}_{ij}$的计算结果

		宗教信仰		
		A	B	C
态度	赞成生育控制	102.22	76.67	51.11
	反对生育控制	97.78	73.33	48.89

4.4.4 列联表的独立性检验

由表4.10 和表4.11 可以计算检验统计量的值

$$\chi^2=\frac{(123-102.22)^2}{102.22}+\frac{(64-76.67)^2}{76.67}+\cdots$$

$$+\frac{(57-48.89)^2}{48.89}=15.59$$

此处$r=2$，$c=3$，$(r-1)(c-1)=2$,若取$\alpha=0.05$，查表有 $\chi^2_{0.99}(2)=9.2103$，由于15.59>9.2103，故拒绝原假设，认为宗教信仰和对待生育控制的态度不是独立的。本例中检验的p 值为0.000。

§ 4.5 正态性检验*

4.5.1 正态概率图

- 正态分布是最常用的分布，用来判断总体分布是否为正态分布的检验方法称为正态性检验，它在实际问题中大量使用。

- 正态概率图可用来作正态性检验，方法如下：利用样本数据在概率纸上描点，用目测方法看这些点是否在一条直线附近，若是的话，可以认为该数据来自正态总体，若明显不在一条直线附近，则认为该数据来自非正态总体。

4.5.1 正态概率图*

- 例4.25 随机选取10个零件，测得其直径与标准尺寸的偏差如下：（单位：丝）

  9.4 8.8 9.6 10.2 10.1 7.2 11.1 8.2 8.6 9.8

在正态概率图上作图步骤如下：

(1) 首先将数据排序：

7.2 8.2 8.6 8.8 9.4 9.6 9.8 10.1 10.2 11.1;

(2) 对每一个i，计算修正频率

$$(i-0.375)/(n+0.25), \quad i=1,2,\ldots,n,$$

4.5.1 正态概率图*

(3) 将点$\left(x_{(i)},\ (i-0.375)/(n+0.25)\right)$，$i=1,2,\cdots,n$ 逐一点在正态概率图上（图4.6），

(4) 观察上述n个点的分布，作出如下判断：

- 若诸点在一条直线附近，则认为该批数据来自正态总体；
- 若诸点明显不在一条直线附近，则认为该批数据的总体不是正态分布。

4.5.1 正态概率图*

从图4.6 可以看到，10个点基本在一条直线附近，故可认为直径与标准尺寸的偏差服从正态分布。

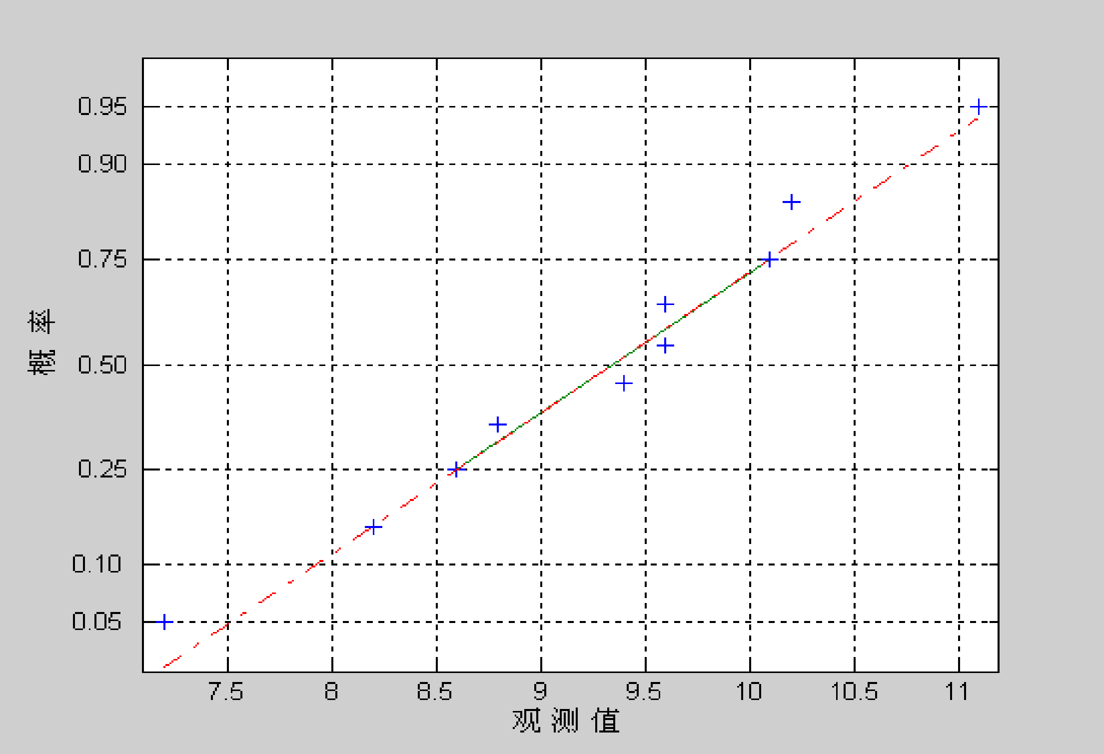


图4.6 例4.25 的正态概率图

4.5.1 正态概率图*

- 如果从正态概率纸上确认总体是非正态分布时，可对原始数据进行变换后再在正态概率纸上描点，若变换后的点在正态概率纸上近似在一条直线附近，则可以认为变换后的数据来自正态分布，这样的变换称为正态性变换。常用的正态性变换有如下三个：对数变换$Y=lnX$、倒数变换$Y=1/X$和根号变换$Y=\sqrt{X}$。

4.5.1 正态概率图*

- 例4.26 随机抽取某种电子元件10个, 测得其寿命数据如下:

  110.47, 99.16, 97.04, 32.62, 2269.82,
  539.35, 179.49, 782.93, 561.10, 286.80.

- 图4.7 给出这10个点在正态概率纸上的图形，这10个点明显不在一条直线附近，所以可以认为该电子元件的寿命的分布不是正态分布。

4.5.1 正态概率图*

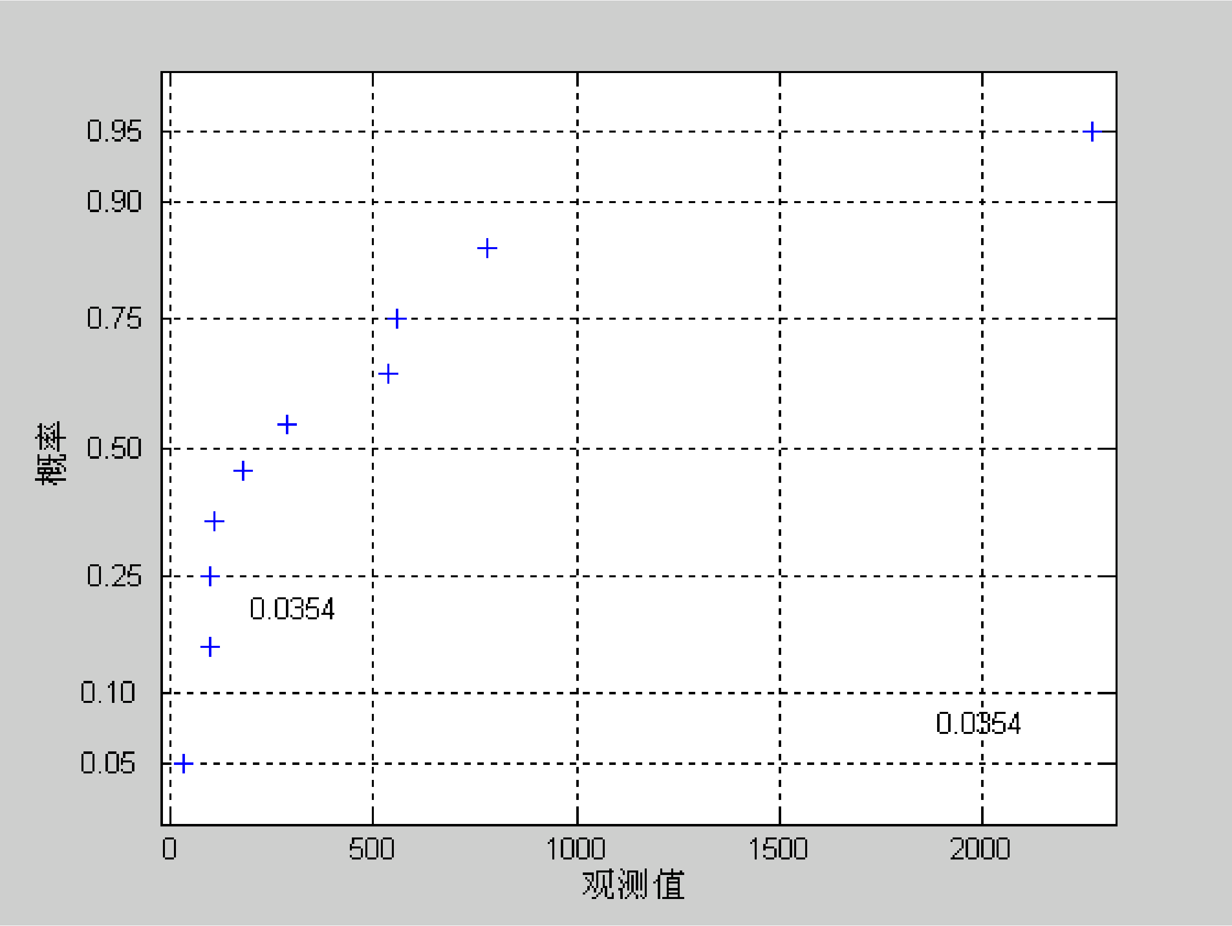


图4.7 例4.26 的正态概率图

4.5.1 正态概率图*

对该10个寿命数据作对数变换, 结果见表4.12。

表4.12　对数变换后的数据

i	$x_{(i)}$	$\ln x_{(i)}$	$\dfrac{i-0.375}{n+0.25}$	i	$x_{(i)}$	$\ln x_{(i)}$	$\dfrac{i-0.375}{n+0.25}$
1	32.62	3.4849	0.061	6	286.80	5.6588	0.549
2	97.04	4.5752	0.159	7	539.35	6.2904	0.646
3	99.16	4.5967	0.256	8	561.10	6.3299	0.743
4	110.47	4.7048	0.354	9	782.93	6.6630	0.841
5	179.49	5.1901	0.451	10	2269.82	7.7275	0.939

4.5.1 正态概率图*

利用表4.12 中最后两列上的数据在正态概率图上描点，结果见图4.8，从图上可以看到10个点近似在一条直线附近，说明对数变换后的数据可以看成来自正态分布。这也意味着，原始数据服从对数正态分布。

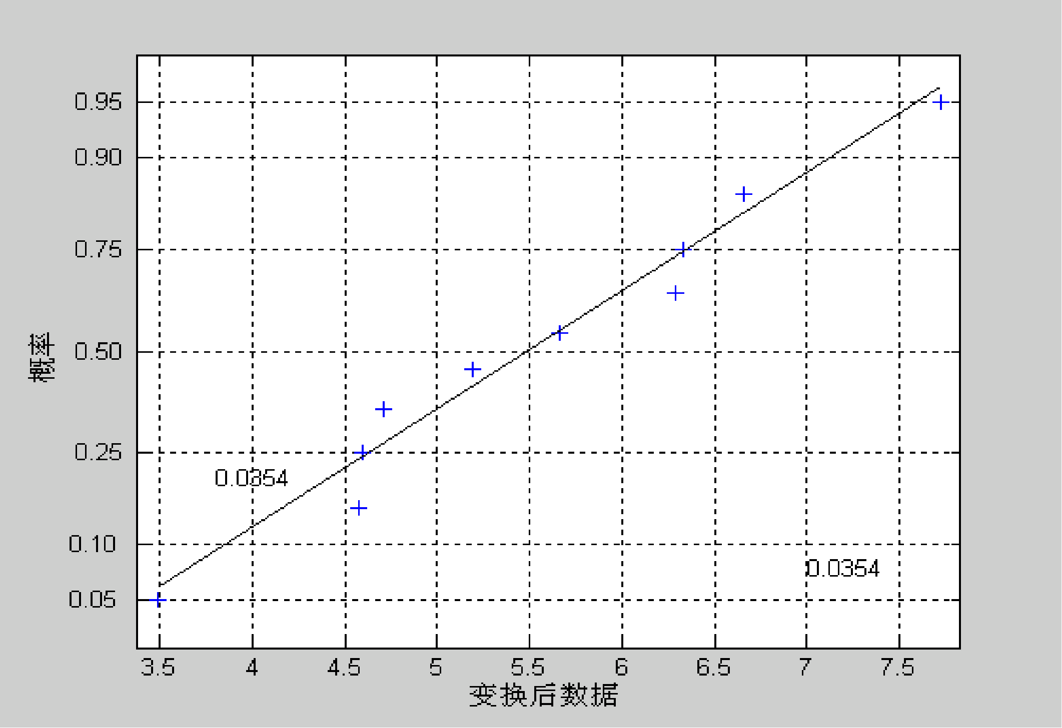


图4.8　变换后数据的正态概率图

4.5.2 *W*检验*

夏皮洛—威尔克检验(Shapiro-Wilk test)也简称*W*检验。这个检验当$8 \le n \le 50$时可以利用。过小样本($n<8$)对偏离正态分布的检验不太有效。

*W*检验是建立在次序统计量的基础上。

检验统计量为：$$W=\frac{\left[\sum_{i=1}^{n}(a_i-\bar{a})(X_{(i)}-\bar{X})\right]^2}{\sum_{i=1}^{n}(a_i-\bar{a})^2\sum_{i=1}^{n}(X_{(i)}-\bar{X})^2} \quad (4.1)$$

其中系数a_i可查附表6。

拒绝域为：$\{W \le W_\alpha\}$，其中α分位数W_α可查附表7。

4.5.2 W检验*

- 系数$a_1, \dots, a_n$还具有如下几条性质：

(1) $a_i = -a_{n+1-i}$， $i = 1,2,\dots,[n/2]$.

(2) $\sum_{i=1}^{n} a_i = 0$.

(3) $\sum_{i=1}^{n} a_i^2 = 1$.

- 据此可将公式(4.1)简化为

$$W = \frac{\left[\sum_{i=1}^{[n/2]} a_i \left(X_{(i)} - X_{(n-i+1)}\right)\right]^2}{\sum_{i=1}^{n}\left(X_{(i)} - \bar{X}\right)^2}$$

4.5.2 *W*检验*

- 例4.27 某气象站收集了44个独立的年降雨量数据，资料如下（已排序）：

520	556	561	616	635	669	686	692	704	707	711
713	714	719	727	735	740	744	745	750	776	777
786	786	791	794	821	822	826	834	837	751	862
873	879	889	900	904	922	926	952	963	1056	1074

我们要根据这批数据作正态性检验。

4.5.2 W检验*

首先由这批数据可算得:

$$\bar{x} = 785.114, \qquad \sum_{i=1}^{44}\left(x_{(i)} - \bar{x}\right)^2 = 630872.43.$$

我们将计算W 的过程列于表4.13中。

为便于计算，值$x_{(k)}$，$x_{(n+1-k)}$和$d_k = x_{(n+1-k)} - x_{(k)}$安排在同一行。

4.5.2 *W*检验*

表4.13 某一气象站收集的年降雨量

k	$x_{(k)}$	$x_{(n+1-k)}$	d_k	a_k
1	520	1074	554	0.3872
2	556	1056	500	0.2667
3	561	963	402	0.2323
4	616	952	336	0.2072
5	635	926	291	0.1868
6	669	922	253	0.1695
7	686	904	218	0.1542

4.5.2 W检验*

续表4.13 某一气象站收集的年降雨量

k	$x_{(k)}$	$x_{(n+1-k)}$	d_k	a_k
8	692	900	208	0.1405
9	704	889	185	0.1278
10	707	879	172	0.1160
11	711	873	162	0.1049
12	713	862	149	0.0943
13	714	851	137	0.0842
14	719	837	118	0.0745
15	727	834	107	0.0651

4.5.2 *W*检验*

续表4.13 某一气象站收集的年降雨量

k	$x_{(k)}$	$x_{(n+1-k)}$	d_k	a_k
16	735	826	91	0.0560
17	740	822	82	0.0471
18	744	821	77	0.0383
19	745	794	49	0.0296
20	750	791	41	0.0211
21	776	786	10	0.0126
22	777	786	9	0.0042

4.5.2 W检验*

从表4.13 可以计算出W的值：

$$W = \frac{(0.3872 \times 554 + 0.2667 \times 500 + \cdots + 0.0042 \times 9)^2}{630872.43}$$
$$= 0.982$$

若取$\alpha = 0.05$，查附表7，在n=44时给出：

$$W_{0.05} = 0.944$$

由于计算得到的W值大于该值，所以在显著性水平$\alpha = 0.05$上不拒绝零假设，即可以认为该批数据服从正态分布。

Thank You !