第八章

假设检验

假设检验的基本概念

若对参数一无所知 → 用参数估计的方法处理

若对参数有所了解 → 但有怀疑猜测需要证实之时 → 用假设检验的方法来处理

§8.1 假设检验的基本思想与概念

假设检验问题

引例　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)。某天从生产中随机抽取25块合金，测得强度值为 $x_1, x_2, \ldots, x_{25}$，其均值为 $\bar{x} = 108$ (Pa)，问当日生产是否正常？

(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：

$$H_0 = \{\theta : \theta \geq 110\} \qquad H_1 = \{\theta : \theta < 110\}$$

这两个非空参数集合都称作统计假设，简称假设。

(4) 我们的任务是利用样本去判断假设（命题）“$\theta \in H_0$”是否成立。这里的“判断”在统计学中称为检验或检验法则。

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，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备择假设，用 H_1 表示，它们常常成对出现。

在上例中，我们可建立如下两个假设：

$$H_0:\theta \geq 110 \quad \text{vs} \quad H_1:\theta < 110$$

二、选择检验统计量，给出拒绝域形式

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。使原假设被拒绝的样本观测值所在区域称为拒绝域，一般用W表示，在上例中，样本均值$\bar{x}$愈大，意味着总体均值θ也大，因此，合理的拒绝域形如

$$W=\{(x_1,\cdots,x_n):\bar{x}\le c\}=\{\bar{x}\le c\}$$

正如在数学上我们不能用一个例子去证明一个结论一样，用一个样本（例子）不能证明一个命题（假设）是成立的，但可以用一个例子（样本）推翻一个命题。因此，从逻辑上看，注重拒绝域是适当的。事实上，在“拒绝原假设”和“拒绝备择假设（从而接收原假设）”之间还有一个模糊域，如今我们把它并入接收域，所以接收域是复杂的，将之称为保留域也许更恰当，但习惯上已把它称为接收域，没有必要再进行改变，只是应注意它的含义。

三、选择显著性水平

由于检验法则是根据样本做出的，所以检验可能犯以下两类错误：

- 其一是 H_0为真但样本观测值落在拒绝域中，从而拒绝原假设 H_0 ，这种错误称为第一类错误，其发生的概率称为犯第一类错误的概率，或称拒真概率，通常记为α.
- 其二是 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, 这种错误称为第二类错误，其发生的概率称为犯第二类错误的概率,或称采伪概率，通常记为 β。

观测数据情况	总体情况	
	H_0为真	H_1为真
$(x_1,\cdots,x_n)\in W$	犯第一类错误	正确
$(x_1,\cdots,x_n)\in W^c$	正确	犯第二类错误

- 当α减小时，必导致β的增大；
- 当β减小时，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的显著性检验的概念。

显著性检验：控制第一类错误，让它不大于 α；而不考虑第二类错误概率。

数 α 称之为显著性水平。

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。

在上例，若取α=0.05,

若令 $u = \dfrac{\overline{x} - 110}{4/5}$

则拒绝域有一种表示：

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

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$u \le -1.645$ 时，则拒绝 H_0
  即接收 H_1 ；
- 当 $u > -1.645$时，则接收 H_0

在上例中，由于 $\overline{x} = 108$代人u中，使得$u < -1.645$.
因此拒绝原假设，即认为该日生产不正常。

知识总结及例题

△ 假设检验的内容

- **参数检验**
  - **总体均值，均值差的检验**
  - **总体方差，方差比的检验**
- **非参数检验**
  - **分布拟合检验**
  - **符号检验**
  - **秩和检验**

假设检验的理论依据

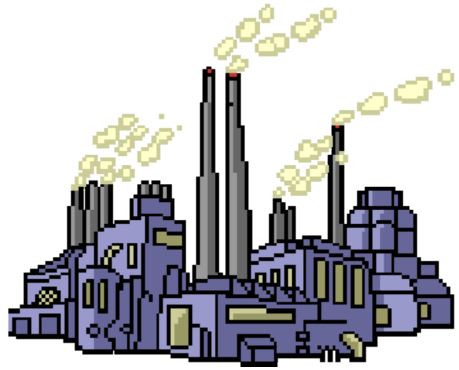
假设检验所以可行,其理论背景为实际推断原理,即"小概率原理"

通常借助于直观分析和理论分析相结合的做法,其基本原理就是人们在实际问题中经常采用的所谓实际推断原理:"一个小概率事件在一次试验中几乎是不可能发生的".

实例　某车间用一台包装机包装葡萄糖, 包得的袋装糖重是一个随机变量, 它服从正态分布. 当机器正常时, 其均值为0.5千克, 标准差为0.015千克. 某日开工后为检验包装机是否正常, 随机地抽取它所包装的糖9袋, 称得净重为(千克):

0.497　0.506　0.518　0.524　0.498　0.511
0.520　0.515　0.512,　问机器是否正常?

分析: 用 μ 和 σ 分别表示这一天袋装糖重总体 X 的均值和标准差,



由长期实践可知, 标准差较稳定, 设 $\sigma = 0.015$,

则 $X \sim N(\mu,\ 0.015^2)$, 其中 μ 未知.

问题: 根据样本值判断 $\mu = 0.5$ 还是 $\mu \neq 0.5$.

提出两个对立假设 $H_0 : \mu = \mu_0 = 0.5$ 和 $H_1 : \mu \neq \mu_0$.

再利用已知样本作出判断是接受假设 H_0 (拒绝假设 H_1) , 还是拒绝假设 H_0 (接受假设 H_1).

如果作出的判断是接受 H_0, 则 $\mu = \mu_0$,

即认为机器工作是正常的, 否则, 认为是不正常的.

由于要检验的假设设计总体均值, 故可借助于样本均值来判断.

因为 $\overline{X}$ 是 μ 的无偏估计量,

所以若 H_0 为真, 则 $|\overline{x}-\mu_0|$ 不应太大,

当H_0为真时, $\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$,

衡量 $|\overline{x}-\mu_0|$ 的大小可归结为衡量 $\dfrac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}}$ 的大小,

于是可以选定一个适当的正数k,

当观察值 x 满足 $\dfrac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}} \geq k$ 时，拒绝假设 H_0，

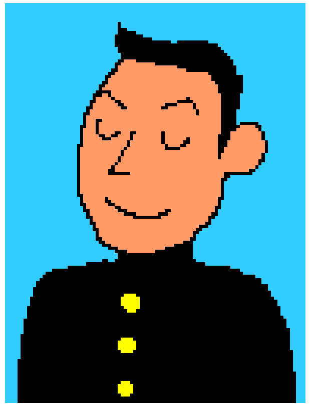
反之，当观察值 x 满足 $\dfrac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}} < k$ 时，接受假设 H_0.

因为当 H_0 为真时 $Z=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$，

由标准正态分布分位点的定义得 $k = z_{\alpha/2}$，

当 $\dfrac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}} \geq z_{\alpha/2}$ 时，拒绝 H_0，$\dfrac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}} < z_{\alpha/2}$ 时，接受 H_0.

假设检验过程如下:



在实例中若取定 $\alpha = 0.05$,

则 $k = z_{\alpha/2} = z_{0.025} = 1.96$,

又已知 $n = 9,\ \sigma = 0.015$,

由样本算得 $\bar{x} = 0.511$, 即有 $\dfrac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} = 2.2 > 1.96$,

于是拒绝假设H_0, 认为包装机工作不正常.

以上所采取的检验法是符合实际推断原理的.

由于通常 α 总是取得很小, 一般取 $\alpha = 0.01, \alpha = 0.05,$

因而当 H_0 为真, 即 $\mu = \mu_0$ 时, $\left\{ \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \geq z_{\alpha/2} \right\}$ 是一个

小概率事件, 根据实际推断原理, 就可以认为如果

H_0 为真, 由一次试验得到满足不等式 $\left| \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \right| \geq z_{\alpha/2}$

的观察值 $\overline{x}$, 几乎是不会发生的.

在一次试验中，得到了满足不等式 $\left|\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq z_{\alpha/2}$ 的观察值 $\bar{x}$，则我们有理由怀疑原来的假设 H_0 的正确性，因而拒绝 H_0.

若出现观察值 $\bar{x}$ 满足不等式 $\left|\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}$，则没有理由拒绝假设 H_0，因而只能接受 H_0.

二、假设检验的相关概念

1. 显著性水平

数 α 称为显著性水平.

上述关于 $\bar{x}$ 与 μ_0 有无显著差异的判断是在显著性水平 α 之下作出的.

2. 检验统计量

统计量 $Z=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 称为检验统计量.

3. 原假设与备择假设

假设检验问题通常叙述为: 在显著性水平α下,

检验假设 $H_0:\mu=\mu_0, H_1:\mu\neq\mu_0.$

或称为“在显著性水平 α下,针对 H_1检验 H_0”.

H_0称为原假设或零假设, H_1 称为备择假设.

4. 拒绝域与临界点

当检验统计量取某个区域C中的值时, 我们拒绝原假设H_0, 则称区域C为**拒绝域**, 拒绝域的边界点称为**临界点**.

如在前面实例中,

拒绝域为 $|z| \geq z_{\alpha/2}$,

临界点为 $z = -z_{\alpha/2}$, $z = z_{\alpha/2}$.

5. 两类错误及记号

假设检验的依据是: 小概率事件在一次试验中很难发生, 但很难发生不等于不发生, 因而假设检验所作出的结论有可能是错误的. 这种错误有两类:

(1) 当原假设H_0为真, 观察值却落入拒绝域, 而作出了拒绝H_0的判断, 称做**第一类错误**, 又叫**弃真错误**, 这类错误是“以真为假”. 犯第一类错误的概率是显著性水平α.

(2) 当原假设 H_0 不真, 而观察值却落入接受域, 而作出了接受 H_0 的判断, 称做**第二类错误**, 又叫**取伪错误**, 这类错误是“以假为真”.

犯第二类错误的概率记为

$$P\{\text{当 } H_0 \text{ 不真接受 } H_0\} \text{ 或 } P_{\mu \in H_1}\{\text{接受 } H_0\}.$$

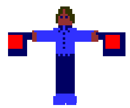
当样本容量 n 一定时, 若减少犯第一类错误的概率, 则犯第二类错误的概率往往增大.

若要使犯两类错误的概率都减小, 除非增加样本容量.

6. 显著性检验

只对犯第一类错误的概率加以控制，而不考虑犯第二类错误的概率的检验，称为**显著性检验**.

7. 双边备择假设与双边假设检验



在 $H_0:\mu=\mu_0$ 和 $H_1:\mu\neq\mu_0$ 中，备择假设 H_1 表示 μ 可能大于 μ_0，也可能小于 μ_0，称为双边备择假设，形如 $H_0:\mu=\mu_0$，$H_1:\mu\neq\mu_0$ 的假设检验称为双边假设检验.

8. 右边检验与左边检验

形如 $H_0: \mu = \mu_0, H_1: \mu > \mu_0$ 的假设检验称为右边检验.

形如 $H_0: \mu = \mu_0, H_1: \mu < \mu_0$ 的假设检验称为左边检验.

右边检验与左边检验统称为单边检验.

9. 单边检验的拒绝域

设总体 $X \sim N(\mu, \sigma^2)$, σ为已知, X_1, X_2, Λ, X_n 是来自总体 X 的样本,给定显著性水平 α,

则 右边检验的拒绝域为 $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \geq z_\alpha$,

左边检验的拒绝域为 $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \leq -z_\alpha$.

证明 (1)右边检验 $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$,

取检验统计量 $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$,

因 H_0 中的全部 μ 都比 H_1 中的 μ 要小,

当 H_1 为真时,观察值 $\overline{x}$ 往往偏大,

因此拒绝域的形式为 $\overline{x} \geq k$, k 为待定正常数,

由 $P\{H_0 \text{为真拒绝} H_0\} = P_{\mu \in H_0}\{\overline{X} \geq k\}$

$$= P_{\mu \le \mu_0}\left\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \geq \frac{k-\mu_0}{\sigma/\sqrt{n}}\right\}$$

$$\leq P_{\mu \le \mu_0}\left\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{k-\mu_0}{\sigma/\sqrt{n}}\right\}$$

上式不等号成立的原因:

因为 $\mu \le \mu_0$, 所以 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \ge \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$,

事件$\left\{\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \ge \dfrac{k-\mu_0}{\sigma/\sqrt{n}}\right\} \subset \left\{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \ge \dfrac{k-\mu_0}{\sigma/\sqrt{n}}\right\}$.

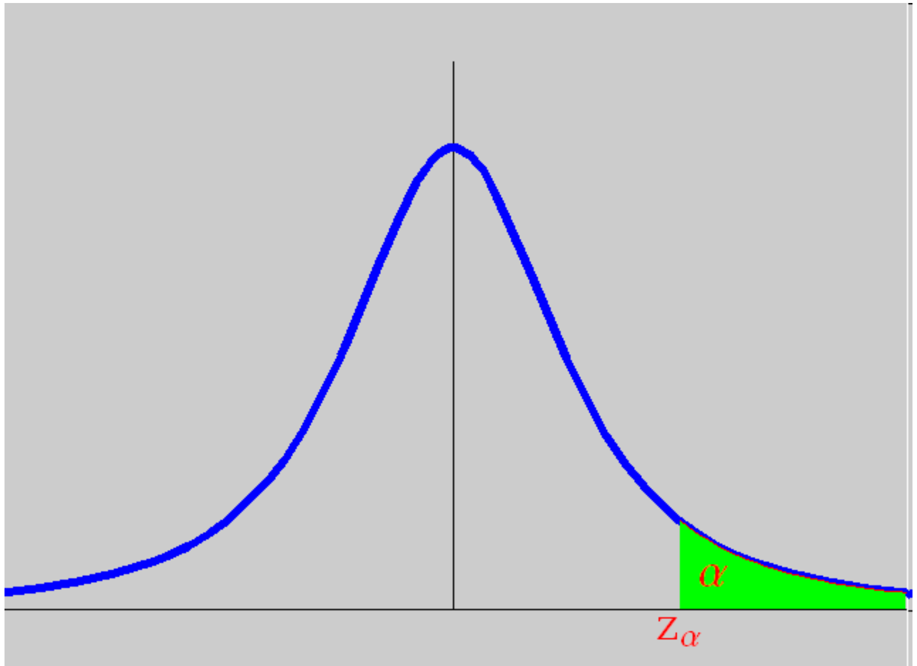
要控制 $P\{H_0$ 为真拒绝 $H_0\} \le \alpha$,

只需令 $P_{\mu \le \mu_0}\left\{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \ge \dfrac{k-\mu_0}{\sigma/\sqrt{n}}\right\} = \alpha$.

因为 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$,

所以 $\dfrac{k-\mu_0}{\sigma/\sqrt{n}} = z_\alpha$,

$$k = \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha,$$



故右边检验的拒绝域为 $\overline{x} \geq \mu_0 + \dfrac{\sigma}{\sqrt{n}} z_\alpha$,

即 $z = \dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$.

证明 (2)左边检验 $H_0: \mu \ge \mu_0$, $H_1: \mu < \mu_0$,

拒绝域的形式为 $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \le k$, k 待定,

由 $P\{H_0 \text{为真拒绝} H_0\} = P_{\mu \ge \mu_0}\left\{\dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \le k\right\} = \alpha$,

得 $k = -z_\alpha$,

故左边检验的拒绝域为 $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \le -z_\alpha$.

三、假设检验的一般步骤

1. 根据实际问题的要求,提出原假设 H_0 及备择假设 H_1;

2. 给定显著性水平 α 以及样本容量 n;

3. 确定检验统计量以及拒绝域形式;

4. 按 $P\{H_0$ 为真拒绝 $H_0\} = \alpha$ 求出拒绝域;

5. 取样,根据样本观察值确定接受还是拒绝H_0.

四、典型例题

例1 某工厂生产的固体燃料推进器的燃烧率服从正态分布 $N(\mu, \sigma^2)$, $\mu = 40\text{cm/s}$, $\sigma = 2\text{cm/s}$. 现用新方法生产了一批推进器, 随机取 $n = 25$只, 测得燃烧率的样本均值为 $\bar{x} = 41.25\text{cm/s}$. 设在新方法下总体均方差仍为 2cm/s,问用新方法生产的推进器的燃烧率是否较以往生产的推进器的燃烧率有显著的提高？取显著水平 $\alpha = 0.05$.

解 根据题意需要检验假设

$H_0: \mu \le \mu_0 = 40$ (即假设新方法没有提高燃烧率),

$H_1: \mu > \mu_0$ (即假设新方法提高了燃烧率),

这是右边检验问题,

拒绝域为 $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \ge z_{0.05} = 1.645$.

因为 $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = 3.125 > 1.645$, z值落在拒绝域中,

故在显著性水平 $\alpha = 0.05$ 下拒绝 H_0.

即认为这批推进器的燃烧率较以往有显著提高.

例2 设 (X_1, X_2, Λ, X_n) 是来自正态总体 $N(\mu, 100)$ 的一个样本，要检验 $H_0: \mu = 0\ (H_1: \mu \neq 0)$，在下列两种情况下，分别确定常数 d，使得以W_1为拒绝域的检验犯第一类错误的概率为0.05 .

(1) $n = 1, W_1 = \{x_1 \,\|\, |x_1| > d\}$;

(2) $n = 25, W_1 = \{(x_1, \Lambda, x_{25}) \,\|\, |\bar{x}| > d\}$，其中 $\bar{x} = \dfrac{1}{25}\sum_{i=1}^{25} x_i$.

解 (1) $n = 1$时，若 H_0 成立，则 $\dfrac{X_1}{10} \sim N(0,1)$,

$$P\left(X_1 \in W_1\right) = P\left(\left|X_1\right| > d\right)$$

$$= P\left(\left|\frac{X_1}{10}\right| > \frac{d}{10}\right) = \Phi\left(-\frac{d}{10}\right) - \Phi\left(\frac{d}{10}\right)$$

$$= 2\left(1 - \Phi\left(\frac{d}{10}\right)\right) = 0.05,$$

$$\Phi\left(\frac{d}{10}\right) = 0.975, \qquad \frac{d}{10} = 1.96, \qquad d = 19.6;$$

(2) $n=25$时，若H_0成立， 则$\sqrt{25}\dfrac{\overline{X}}{10}\sim N(0,1)$，

$$P\left(\left(X_1,\Lambda\ X_{25}\right)\in W_1\right)=P\left(\left|\overline{X}\right|>d\right)$$

$$=P\left(\left|\frac{\overline{X}}{2}\right|>\frac{d}{2}\right)=\Phi\left(-\frac{d}{2}\right)-\Phi\left(\frac{d}{2}\right)$$

$$=2\left(1-\Phi\left(\frac{d}{2}\right)\right)=0.05,$$

$$\Phi\left(\frac{d}{2}\right)=0.975,\qquad \frac{d}{2}=1.96,\qquad d=3.92\,.$$

例3　设 (X_1, X_2, Λ, X_n) 是来自正态总体 $N(\mu, 9)$ 的一个样本,其中 μ 为未知参数,检验 $H_0: \mu = \mu_0$ $(H_1: \mu \neq \mu_0)$,拒绝域 $W_1 = \{(x_1, \Lambda, x_n) \mid |\bar{x} - \mu_0| \geq C\}$,

(1) 确定常数 C,使显著性水平为 0.05;

(2) 在固定样本容量 $n = 25$ 的情况下,分析犯两类错误的概率 α 和 β 之间的关系.

解　(1) 若 H_0 成立，则 $\dfrac{\sqrt{n}(\bar{X} - \mu_0)}{3} \sim N(0,1)$,

$$P((X_1, \Lambda \ X_n) \in W_1) = P(\left|\bar{X} - \mu_0\right| \geq C)$$

$$= P\left(\frac{\sqrt{n}\left|\overline{X}-\mu_0\right|}{3} \geq \frac{\sqrt{n}C}{3}\right) = 2\left(1-\Phi\left(\frac{\sqrt{n}C}{3}\right)\right) = 0.05,$$

$$\Phi\left(\frac{\sqrt{n}C}{3}\right) = 0.975, \qquad \frac{\sqrt{n}C}{3} = 1.96, \qquad C = \frac{5.88}{\sqrt{n}};$$

(2) $n = 25$时, 若H_0成立,

$$\alpha = P((X_1, \Lambda\ X_n) \in W_1)$$

$$= 2\left(1-\Phi\left(\frac{\sqrt{n}C}{3}\right)\right) = 2\left(1-\Phi\left(\frac{5C}{3}\right)\right).$$

若H_0不成立，不妨假设 $\mu = \mu_1 \neq \mu_0$，

$$\beta = P((X_1, \Lambda\ X_n) \overline{\in} W_1) = P(|\overline{X} - \mu_0| < C)$$

$$= P(-C + \mu_0 < \overline{X} < C + \mu_0)$$

$$= P\left(\frac{5}{3}(-C + \mu_0 - \mu) < \frac{5}{3}(\overline{X} - \mu) < \frac{5}{3}(C + \mu_0 - \mu)\right)$$

$$= \Phi\left(\frac{5}{3}(C + \mu_0 - \mu_1)\right) - \Phi\left(\frac{5}{3}(-C + \mu_0 - \mu_1)\right),$$

当C 较小时，α 较大，β 较小；

当C 较大时，α 较小，β 较大.

由于 μ 是任取的，所以对所有 $\mu \neq \mu_0$ 上面的关系成立.

五

假设检验的基本原理、相关概念和一般步骤.

假设检验的两类错误

真实情况(未知)	所作决策	
	接受 H_0	拒绝 H_0
H_0 为真	正确	犯第I类错误
H_0 不真	犯第II类错误	正确