3.3、检验的p值和最优检验的概念

一、　检验的p 值

假设检验的结论通常是简单的：在给定的显著水平下，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：

1

因为显著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？

例：：一支香烟中的尼古丁含量X 服从正态分布$N(\mu, 1)$，质量标准μ 规定不能超过1.5毫克。现从某厂生产的香烟中随机抽取20支测得其中平均每支香烟的尼古丁含量为1.97毫克，试问该厂生产的香烟尼古丁含量是否符合质量标准的规定。

这是一个假设检验问题：

H_0：$\mu \le 1.5$,　H_1：$\mu > 1.5$,

采用u检验，计算得：

$$u = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{1.97 - 1.5}{1/\sqrt{20}} = 2.10$$

对一些的显著性水平，下表列出了相应的拒绝域和检验结论。

显著性水平	拒绝域	u=2.10对应的结论
α =0.05	$u \ge 1.645$	拒绝H_0
α =0.025	$u \ge 1.96$	拒绝H_0
α =0.01	$u \ge 2.33$	接受H_0
α =0.005	$u \ge 2.58$	接受H_0

我们看到，不同的α 有不同的结论。

现在换一个角度来看，在μ =1.5时，u的分布是$N(0,1)$。此时可算得，$P(u \ge 2.10)=0.0179$，若以0.0179为基准来看上述检验问题，可得

- 当α <0.0179时，u >2.10。于是2.10就不在$\{u \ge z_{1-\alpha}\}$ 中，此时应接受原假设H_0；
- 当$\alpha \ge 0.0179$时，u ≤2.10。于是2.10就落在$\{u \ge z_{1-\alpha}\}$中，此时应拒绝H_0。

由此可以看出，0.0179是能用观测值2.10做出"拒绝H_0"的最小的显著性水平，这就是p值。

定义：　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下保留 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。

下面给出一些常见形式的拒绝域所对应的 p 值计算式.

设检验统计量 T 为连续型变量，T_α 表示 T 分布的下侧分位数，检验的显著性水平为 α.

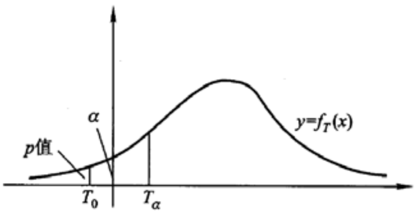
情形 1，若检验的拒绝域为 $T \leqslant T_\alpha$,

即有 H_0 为真时 $P(T \leqslant T_\alpha) = \alpha$.

检验统计量的值为 T_0，则该检验的 p 值计算式为

$$p = P(T \leqslant T_0)\ (H_0\text{ 为真条件下});$$

这时 $T_0 \leqslant T_\alpha$ 成立与 $p \leqslant \alpha$ 成立等价.



p 值示意图

情形 2，若检验的拒绝域为 $T > T_{1-\alpha}$，检验统计量的值为 T_0，则该检验的 p 值计算式为

$$p = P(T > T_0)\ (H_0\text{ 为真条件下});$$

情形 3，若检验的拒绝域为 $|T| > T_{1-\frac{\alpha}{2}}$，检验统计量的值为 T_0，则该检验的 p 值计算式为

$$p = P(|T| > |T_0|)\ (H_0\text{ 为真条件下}).$$

例 单个正态总体均值的检验问题

设 $X \sim N(\mu, 9)$，从 X 得到容量为 10 的样本均值为 $\bar{X} = 2.98$，检验假设 $H_0: \mu = 1$，检验水平 $\alpha = 0.05$，计算检验的 p 值.

解 检验的拒绝域为

$$|U| = \left|\frac{\bar{X}-1}{3}\sqrt{10}\right| > z_{0.975} = 1.96,$$

因为

$$|U_0| = \left|\frac{2.98-1}{3}\sqrt{10}\right| = 2.09,$$

所以检验的 p 值

$$\begin{aligned} p &= P(|U| > |U_0|) = P\left(\left|\frac{\bar{X}-1}{3}\sqrt{10}\right| > 2.09\right) \\ &= 1 - P\left(\left|\frac{\bar{X}-1}{3}\sqrt{10}\right| \leqslant 2.09\right) \\ &= 1 - \Phi(2.09) + \Phi(-2.09) \\ &= 2 - 2 \times 0.9817 = 0.0366. \end{aligned}$$

由 p 值的意义可知当显著性水平 α 降低到 0.036 6 时仍会做出拒绝的选择.

p 值可称为观测到的显著性水平.

二、检验结果的实际意义

在假设检验中，原假设 H_0 与备择假设 H_1 虽是互补的，但地位是不对等的．

一般来说检验水平 α 是较小的，因而检验推断是"保护"原假设，而"歧视"备择假设的．

为更好理解 H_0 与 H_1 地位的不对等性，举例如下．

例 设总体 $X\sim N(\mu,1)$，样本均值 $\bar{x}=0.3$，样本容量 $n=1$，取 $\alpha=0.05$，欲检验 $\mu_0=0$，还是 $\mu_1=1$.

这里有两种假设提法，列出如下．

(1) H_0：$\mu=\mu_0$；H_1：$\mu=\mu_1$

(2) H_0：$\mu=\mu_1$；H_1：$\mu=\mu_0$

对于 (1)，应取拒绝域 $W_1=\{U\geqslant u_{0.95}=1.645\}$，

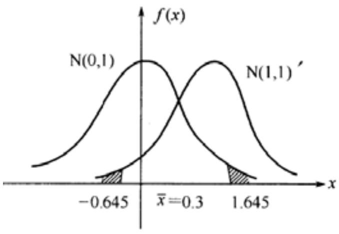
其中 $U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$，当 H_0 成立时，$U=\overline{X}\sim N(0,1)$，

由于 $u=0.3<1.645$，故接受 H_0，即认为 $\mu=\mu_0=0$.

对于 (2)，应取拒绝域 $W_2=\{U\leqslant u_{0.05}=-1.645\}$，

其中 $U=\dfrac{\overline{X}-\mu_1}{\sigma/\sqrt{n}}$，由于 $u=\dfrac{0.3-1}{1/\sqrt{1}}=-0.7>-1.645$，

故接受 H_0，即认为 $\mu=\mu_1=1$.



三、两类错误和犯两类错误的概率、最佳检验的概念

对假设 H_0：$\theta\in\Theta_0$；H_1：$\theta\in\Theta_0^c$ 的检验有可能犯错误

假设检验中的两类错误

假设 \ 决策	拒绝 H_0	接受 H_0
H_0 为真	第一类错误	正确
H_1 为真	正确	第二类错误

记检验的拒绝域为 W，则接受域为 $\overline{W}$；

记样本 $X=(X_1,X_2,\cdots,X_n)$.

若 $\theta\in\Theta_0$，则犯第一类错误的概率为

$$\alpha(\theta)\triangleq P_\theta\{X\in W\mid\theta\}；$$

若 $\theta\in\Theta_0^c$，则犯第二类错误的概率为

$$\beta(\theta)=P_\theta\{X\in\overline{W}\mid\theta\}=1-P_\theta\{X\in W\mid\theta\}.$$

可见两类错误的概率是依赖于 θ 和 W 的二元函数．

记 $P_\theta(X\in W)=P_\theta\{X\in W\mid\theta\}$，定义

$$g(\theta)=P_\theta(X\in W)=\begin{cases}\alpha(\theta), & \theta\in\Theta_0\\ 1-\beta(\theta), & \theta\in\Theta_0^c\end{cases}$$

为具有拒绝域 W 的势函数．

若给定常数 $\alpha(0<\alpha<1)$，$P_\theta(X\in W)\leqslant\alpha$，$\theta\in\Theta_0$

则称 α 为拒绝域 W 的检验水平．

势函数的意义在于当 H_0 不真时，反映拒绝 H_0 的功效大小．

因此，寻找 H_0 的最佳检验法，可归结为选择这样的拒绝域 W，当 H_0 不真时，使拒绝 H_0 的功效最大，即犯第二类错误的概率最小．

一般地当参数空间 $\Theta=\{\theta_0,\theta_1\}$，$\Theta_0=\{\theta_0\}$，$\Theta_0^c=\{\theta_1\}$，那么势函数

$$g(\theta)=\begin{cases}\alpha, & \theta=\theta_0\\ 1-\beta, & \theta=\theta_1\end{cases}$$

例 设总体 $X \sim b(1,\theta)$，$(X_1, X_2, \cdots, X_5)$ 是样本．考虑检验

H_0：$\theta \leqslant \frac{1}{2}$；$H_1$：$\theta > \frac{1}{2}$.

第一种检验法：当且仅当 5 次重复试验中均“成功”时，即 $\sum_{i=1}^{5} x_i = 5$ 时，拒绝 H_0.

此检验法的势函数：$g_1(\theta) = P_\theta\left(\sum_{i=1}^{5} x_i = 5\right) = \theta^5$.

当 $\theta \leqslant \frac{1}{2}$，计算得犯第一类错误的概率

$$\alpha(\theta) = g_1(\theta) \leqslant \left(\frac{1}{2}\right)^5 = 0.0312,$$

当 $\theta > \frac{1}{2}$，犯第二类错误的概率 $\beta(\theta)$ 太大（$g_1(\theta)$ 太小），

当且仅当 $\theta > \left(\frac{1}{2}\right)^{1/5} = 0.87$ 时，$\beta(\theta) \leqslant 0.5$.

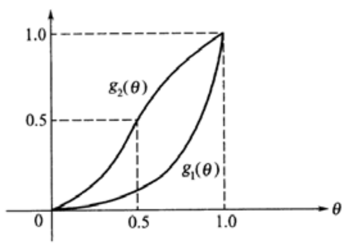
第二种检验法：当 $\sum_{i=1}^{5} x_i = 3$，4 或 5 时，拒绝 H_0.

此检验法的势函数：$g_2(\theta) = P_\theta\left(\sum_{i=1}^{5} x_i = 3, 4 \text{ 或 } 5\right)$

$$= \binom{5}{3}\theta^3(1-\theta)^2 + \binom{5}{4}\theta^4(1-\theta) + \binom{5}{5}\theta^5$$

对于 $\theta > \frac{1}{2}$，有较小的 第二类错误的概率 $\beta(\theta) = 1 - g_2(\theta)$，

但对于 $\theta \leqslant \frac{1}{2}$，却有较大的第一类错误的概率



势函数

例 设 $(X_1, X_2, \cdots, X_n)$ 是来自总体 $N(\mu, \sigma^2)$ 的样本，σ^2 已知，求对假设 H_0：$\mu = \mu_0$；H_1：$\mu = \mu_1 > \mu_0$ 的 u 检验的势函数及两类错误的概率．

解 在此检验中，拒绝域为 $(\overline{X} - \mu_0)/(\sigma/\sqrt{n}) \geqslant c$，$c$ 为大于零的任意常数．此检验的势函数

$$g(\mu) = P_\mu\left\{\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geqslant c\right\} = P_\mu\left\{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \geqslant c + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right\}$$

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

当取检验水平为 α 时，$c = u_{1-\alpha}$，故

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

于是犯第一类错误的概率为：

$$g(\mu_0) = 1 - \Phi(u_{1-\alpha}) = \alpha,$$

犯第二类错误的 概率为

$$\beta = 1 - g(\mu_1) = \Phi\left(u_{1-\alpha} - \frac{\mu_1 - \mu_0}{\sigma}\sqrt{n}\right)$$

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

(1) 设 α、β 均已确定，求样本容量 n.

(2) 设 $\alpha = 0.05$，当 $\mu_1 = \mu_0 + 0.5\sigma$，欲使 $\beta \leqslant 0.1$，n 应取多大?

解 (1) $\beta = \Phi\left(u_{1-\alpha} - \frac{\mu_1 - \mu_0}{\sigma}\sqrt{n}\right)$，

$u_\beta = u_{1-\alpha} - \frac{\mu_1 - \mu_0}{\sigma}\sqrt{n}$，即 $n = \left[\frac{\sigma(u_{1-\alpha} - u_\beta)}{\mu_1 - \mu_0}\right]^2$

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

$0.05 = g(\mu_0) = 1 - \Phi(u_{0.95})$ 故 $u_{0.95} = 1.645$.

$\beta = 1 - g(\mu_0 + \sigma) = \Phi(1.645 - 0.5\sqrt{n}) \leqslant 0.1$，

$0.5\sqrt{n} - 1.645 \geqslant 1.28$，解得 $n \geqslant 34.22$，

取 $n = 35$ 时，能在 $\alpha = 0.05$ 下，使 $\beta \leqslant 0.1$.

最佳检验的概念

显著性检验只注重考虑控制犯第一类错误的概率.

Neyman 和 Pearson 提出了综合考虑控制犯两类错误概率的假设检验方法,它是一种较完善的假设检验方法.

其基本思想是:在控制犯第一类错误的概率条件下,使犯第二类错误的概率达到最小.

设原假设和备择假设分别为 $H_0: \theta \in \Theta_0; H_1: \theta \in \Theta_1$,

检验统计量用 T 表示,

用统计量划分的接受域和拒绝域分别用 $\Re_0, \Re_1$ 表示,

犯第一类错误的概率和犯第二类错误的概率分别用 α, β 表示,

在 $P\{T \in \Re_1\} \leqslant \alpha, \theta \in \Theta_0$ 条件下,找使 $P\{T \in \Re_1\}$尽可能大(即接近 α)的 检验. 从而使犯第二类错误的概率达到最小.

即找使下式成立的检验:

$$\sup_{\theta \in \Theta_0} P\{T \in \Re_1\} \leqslant \alpha.$$

例 单个正态总体均值的检验问题

设 $X \sim N(\mu, \sigma^2)$,σ^2 已知,从 X 得到的样本为 $X_1, X_2, \cdots, X_n$,检验假设 H_0: $\mu \geqslant \mu_0$.

解 拒绝域形式为 $\bar{X} < c$,

$$\sup_{\mu \geqslant \mu_0} P(\bar{X} < c) = \sup_{\mu \geqslant \mu_0} \Phi\left(\frac{c-\mu}{\sigma}\sqrt{n}\right) = \Phi\left(\frac{c-\mu_0}{\sigma}\sqrt{n}\right) = \alpha,$$

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

所以拒绝域为 $\bar{X} < \mu_0 + z_\alpha \dfrac{\sigma}{\sqrt{n}}$,即$\dfrac{\bar{X}-\mu_0}{\sigma}\sqrt{n} < z_\alpha$,

与显著性检验得到的一致.

例 单个正态总体方差的检验问题

设 $X \sim N(\mu, \sigma^2)$,从 X 得到的样本为 $X_1, X_2, \cdots, X_n$,检验假设 H_0: $\sigma^2 \leqslant \sigma_0^2$.

解 拒绝域形式为 $S^2 > c$,

$$\sup_{\sigma^2 \leqslant \sigma_0^2} P(S^2 > c) = \sup_{\sigma^2 \leqslant \sigma_0^2} P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)c}{\sigma^2}\right)$$

$$= \sup_{\sigma^2 \leqslant \sigma_0^2}\left[1 - F_{\chi^2(n-1)}\left(\frac{(n-1)c}{\sigma^2}\right)\right]$$

$$= 1 - F_{\chi^2(n-1)}\left(\frac{(n-1)c}{\sigma_0^2}\right) = \alpha,$$

所以$\dfrac{(n-1)c}{\sigma_0^2} = \chi_\alpha^2(n-1)$,即 $c = \chi_\alpha^2(n-1)\dfrac{\sigma_0^2}{(n-1)}$,

所以拒绝域为

$$S^2 > \chi_\alpha^2(n-1)\frac{\sigma_0^2}{(n-1)},$$

即$\dfrac{(n-1)S^2}{\sigma_0^2} > \chi_\alpha^2(n-1)$,与显著性检验得到的一致.

拒绝域形式给定的条件下局部最佳的

定义

对假设检验问题 $H_0: \theta \in \Theta_0; H_1: \theta \in \Theta_1$,记样本为 $\tilde{X} = (X_1, X_2, \cdots, X_n)'$,检验水平为 α,考察样本的取值范围的划分,取定的一种划分对应的接受域和拒绝域分别用 $\Re_0^*, \Re_1^*$ 表示,任意的一种划分对应的接受域和拒绝域分别用 $\Re_0, \Re_1$ 表示,满足下列两个条件:

(1) $\sup\limits_{\theta \in \Theta_0} P\{\tilde{X} \in \Re_1^*\} \leqslant \alpha, \sup\limits_{\theta \in \Theta_0} P\{\tilde{X} \in \Re_1\} \leqslant \alpha$;

(2) $P\{\tilde{X} \in \Re_0^*\} \leqslant P\{\tilde{X} \in \Re_0\}, \theta \in \Theta_1$.

则称 $\Re_1^*$ 为检验水平为 α 的最佳拒绝域,对应的检验为最佳检验.

作业

P90：14,15,16