假设检验

- 参数假设检验的原理
- 参数假设检验的步骤
- 原假设与备择假设的选取原则
- 假设检验的 p 值

假设检验的概念

假设(*hypothesis*)是关于总体的概率分布或参数的某个命题.

怎么判断所作的假设是否正确?

从总体中抽取样本, 根据样本的取值, 按一定原则进行检验, 然后作出接受或拒绝所作假设的决定, 称为假设检验(hypothesis testing).

假设检验
- 参数检验 —— 总体分布已知,关于未知参数的检验假设
- 非参数检验 —— 总体分布未知,对分布类型的假设检验

- 若对参数一无所知：用参数估计
- 若对参数有所了解，但不确定，需要证实：用参数检验

非参数检验

例 在 50 min 内记录每15s公路上通过汽车的车辆数, 得如下数据， 问能否认为每15s通过的车辆数服从泊松分布?

检验假设“总体服从泊松分布” 是否成立.

车辆数	0	1	2	3	4	5
频数	92	68	28	11	1	0

参数检验

例 某工厂在正常情况下, 生产的灯泡使用寿命 X (单位: h) 服从正态分布 $N\left(5800, 100^2\right)$. 今从该厂生产的一批灯泡中随机地取 25 个进行检测, 测得其平均寿命 $\bar{x} = 5756.6$. 如果标准差不变, 能否认为该厂生产的这批灯泡的寿命均值为 5800 h ?

已知总体 (灯泡寿命) 的分布形式 (正态分布), 要检验假设“总体均值为5800h ” 是否成立.

例：有一个盒子，装有100个白球和红球。两种球的比例是99:1。问这个盒子中是 99个白球1个红球，还是 99个红球1个白球?

假设：这个盒子里有99个白球1个红球.

检验：从中随机摸出一个球，发现是

假设成立时，摸出红球的概率只有1/100，这是小概率事件.

小概率原理(小概率事件在一次试验中基本上不会发生) 。

小概率事件在一次试验中竟然发生了，就应怀疑所作的假设.

这种推理方法，可以称为**带概率性质的反证法**,或称为**概率反证法.**

- 一般反证法是在假定某个假设成立，若导出了与事实相矛盾的结论，则完全绝对地否定原假设.
- 概率反证法是假定某个假设成立，若推出小概率事件发生，则以很大的把握否定原假设.

例 某工厂生产的一种螺钉，标准要求长度是32.5毫米. 实际生产的产品长度X假定服从正态分布$N(\mu,9)$，现从该厂生产的一批产品中抽取16件，平均长度是 $\overline{X}$=30.5，问这批产品是否合格?

问题描述为:已知总体$X \sim N(\mu, 9)$, $\overline{X} = 30.5$

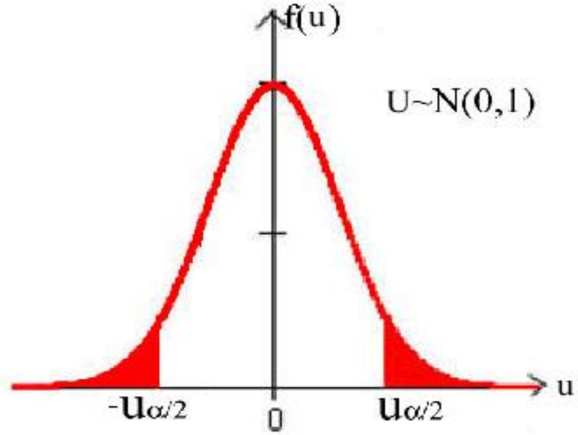
要检验假设 H_0: $\mu = \mu_0 = 32.5$;　H_1:$\mu \neq 32.5$ 哪个成立?

通常称H_0为原假设(Null Hypothesis)，H_1为备择假设(或对立假设，Alternative Hypothesis).

定义检验统计量(test statistic)　$U = \dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

当H_0成立时, $U \sim N(0,1)$，U的值应以大概率集中于“0”的周围;

当$|U| >>0$时，意味着小概率事件发生了, 应该拒绝H_0;



H_0成立时，取常数C 使$|U|>C$为小概率事件（概率为α），

即 $P(|U|>C \mid H_0\text{成立})=\alpha$.　α应当很小，这是拒绝H_0的概率。

H_0的拒绝域(rejection region，critical regions)为：

$$V=\{(x_1,x_2,\ldots,x_n):\ |U|>C\}$$

临界值
(Critical value)

$\bar{V}$称为接受域(acceptance region)

这种假设检验通常称为显著性假设检验(null hypothesis significance testing).

小概率α称为检验水平或称显著性水平(significance level).

α 越大，拒绝域越大，对H_0为真要求越严格。

α 的选择要根据实际情况而定，常取$\alpha=0.1, 0.05, 0.01$.

例题中当H_0成立时, $U \sim N(0,1)$, 因此$C = u_{\alpha/2}$.

若取α=0.05，则$u_{\alpha/2} = u_{0.025} = 1.96$.

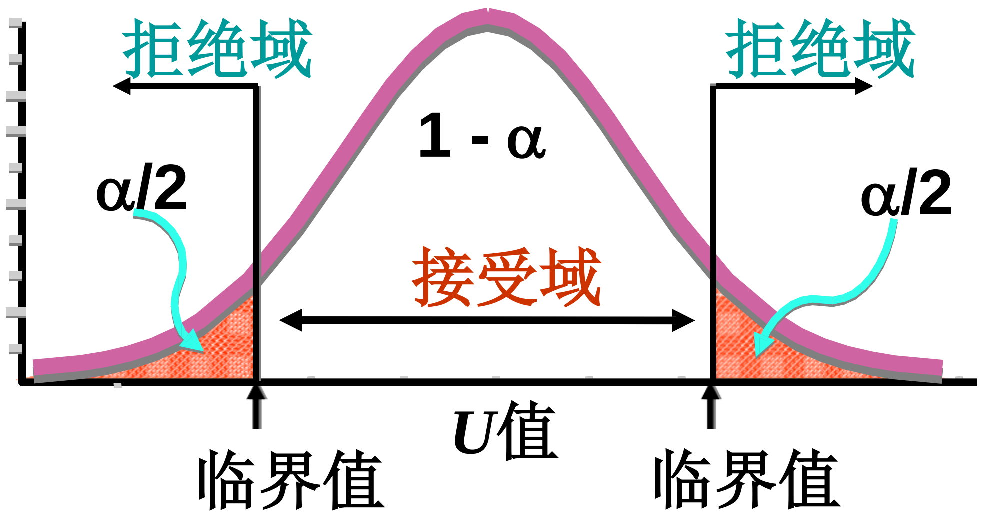
$$V = \{(x_1, x_2, \cdots, x_n):\ |U| > 1.96\}$$

将样本观测值 $\bar{X} = 30.5$ 代入得

$$|U| = \left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| = \left|\frac{30.5-32.5}{3/\sqrt{16}}\right| = 2.67 > 1.96$$

这表明样本落在拒绝域V内，对H_0不利的小概率事件发生了，因此应**拒绝**H_0.

> **说明：** 接受(拒绝)H_0并不是肯定H_0一定对(错)，而只是说现在的样本数据还不能否定(肯定)H_0.
>
> retain(~~accept,~~ fail to reject) / reject H_0



假设检验的步骤

(1) 提出原假设H_0和备择假设H_1(根据题意)

例子中 $H_0: \mu = 32.5 \Leftrightarrow H_1: \mu \neq 32.5$

(2) 选取检验统计量，在H_0成立下确定其分布(已知分布类型)

例子中 $U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

(3) 对给定的显著性水平α，寻找对H_0不利的小概率事件，确定拒绝域V.

例子中$\alpha = 0.05$, $V = \{(x_1, x_2, \cdots, x_n): |U| > 1.96\}$.

(4) 计算检验统计量的值，判断是否拒绝H_0.

例子中 $|U| = 2.67 > 1.96$，落入V内，拒绝H_0.

参数假设检验可写成如下的统一形式

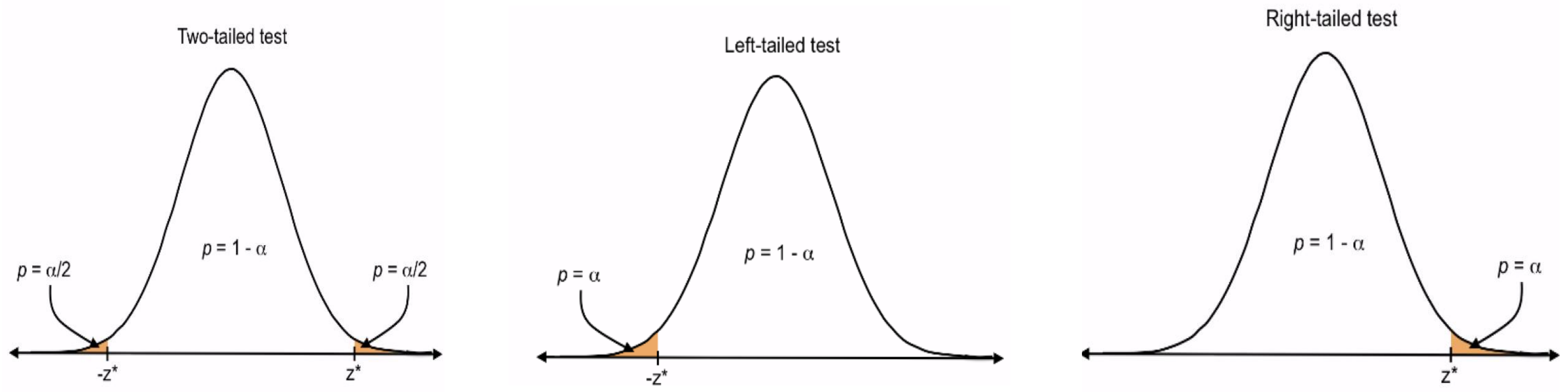
$$H_0: \theta \in \Theta_0 \quad \Leftrightarrow \quad H_1: \theta \in \Theta_1$$

其中Θ是参数空间，Θ_0与Θ_1非空，最常见 $\Theta_1 = \Theta - \Theta_0$

- 如果Θ_1位于Θ_0的两侧，这样的检验称为双侧检验(two-sided)；
- 如果Θ_1位于Θ_0的右侧，这样的检验称为右侧检验(right-tailed)；
- 如果Θ_1位于Θ_0的左侧，这样的检验称为左侧检验(left-tailed).

右侧检验与左侧检验统称为单侧检验(one-sided).

只有1个值($\theta = \theta_0$)的假设称为简单假设(simple)，否则称为复合假设(composite).



原假设与备择假设的选取

假设检验的原假设$\boldsymbol{H_0}$和备择假设$\boldsymbol{H_1}$取决于研究者对问题的态度。

通常原假设不能轻易被否定，处于“被保护”的地位.

一般说来，按以下原则确定：

(1)把研究者要证明的结论作为备择假设$\boldsymbol{H_1}$;

(2)把研究者要反对的假设作为原假设$\boldsymbol{H_0}$;

(3)把现状作为原假设$\boldsymbol{H_0}$;（no change）

(4)把不能轻易否定的假设作为原假设$\boldsymbol{H_0}$.

例如：

为了检验生活在平原地区和高原地区的人的肺活量是否相同，应将肺活量相同作为原假设.

某厂家声称其产品优等品率为70%，在对这一声明进行假设检验时，应将优等品率为70％作为原假设.

填空题 1分

设置

为了研究男性长跑运动员的心率是否低于一般健康男性心率,一医生从某省长跑队随机抽取了25名运动员，测得其平均值为60次/分，标准差为6次/分。大量的资料显示一般健康男性平均心跳为72次/分。假设心率的分布服从正态分布,均值为μ, 是否有理由认为男性长跑运动员每分钟的心跳次数较一般健康男性少?

本题的原假设与备择假设分别为：H_0: μ=72, H_1: μ<72.

A 正确 B 错误 [填空1] (填A或B)

进行假设检验时，使用 **z检验公式** 来计算标准化检验统计量：

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

变量说明：

- $\bar{x} = 60$: 样本均值
- $\mu_0 = 72$: 原假设下的总体均值
- $\sigma = 6$: 样本标准差（假设已知）
- $n = 25$: 样本量

代入数据计算：

$$z = \frac{60 - 72}{\frac{6}{\sqrt{25}}} = \frac{-12}{\frac{6}{5}} = \frac{-12}{1.2} = -10$$

检验结果：

1. 查标准正态分布表，显著性水平为 $\alpha = 0.05$ 的单侧检验临界值为 -1.645。
2. $z = -10$ 显著小于 -1.645。

结论：

拒绝原假设 $H_0 : \mu = 72$，有足够证据支持备择假设 $H_1 : \mu < 72$。

作答

例 某批发商欲从生产厂家购进一批灯泡, 根据合同规定，灯泡的平均使用寿命不能低于3000小时. 已知灯泡使用寿命服从正态分布，标准差为80小时. 今在该批灯泡中随机抽取16只灯泡，测得样本均值为2960小时. 假定标准差不变，问批发商是否应该购买这批灯泡? (α=0.05)

解：根据题意，原假设和备择假设如下：

$$H_0: \mu \geq \mu_0 = 3000, H_1: \mu < 3000$$ ——左侧检验

取 $$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 3000}{80/\sqrt{16}} = \frac{\bar{X} - 3000}{20}$$

作为检验统计量。

$$U=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}+\frac{\mu-\mu_0}{\sigma/\sqrt{n}},\qquad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)\qquad 故\ U\sim N\left(\frac{\mu-\mu_0}{\sigma/\sqrt{n}},1\right)$$

$$\boldsymbol{H_0:\mu\geq\mu_0\,,\,H_1:\mu<\mu_0}$$

当$\boldsymbol{H_0}$成立时，U的均值是正数，此时U的值集中在“正数”的周围

当$\boldsymbol{H_1}$成立时，U的均值是非正数，此时U的值集中在“非正数”的周围

因此，当U的值偏小时，对原假设$\boldsymbol{H_0}$不利，所以拒绝域为

$$V=\{(x_1,x_2,\cdots,x_n):U<C\}$$

$$P(U<C\mid H_0)=P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}<C\mid\mu\geq\mu_0\right)=P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}+\frac{\mu-\mu_0}{\sigma/\sqrt{n}}<C\mid\mu\geq\mu_0\right)$$

$$=P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<C-\frac{\mu-\mu_0}{\sigma/\sqrt{n}}\mid\mu\geq\mu_0\right)\leqslant P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<C\mid\mu\geq\mu_0\right)=\boldsymbol{\alpha}$$

取 $C=\boldsymbol{-u_\alpha}$ $\qquad P(U<\boldsymbol{-u_\alpha}\mid H_0)\leqslant P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\boldsymbol{-u_\alpha}\mid\mu\geq\mu_0\right)=\boldsymbol{\alpha}$

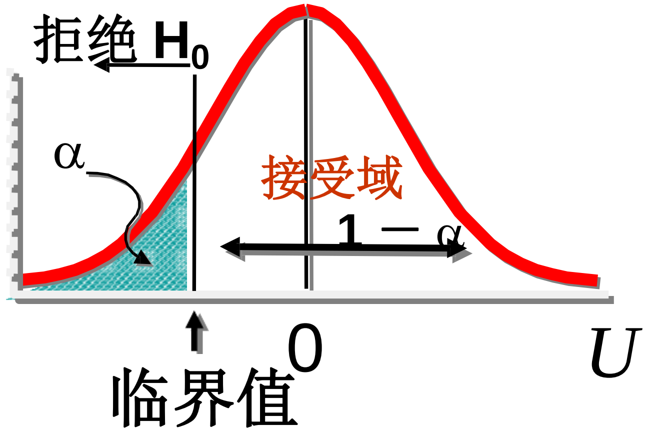
左侧检验：$H_0: \mu \geq 3000,\ H_1: \mu < 3000$

给定α, H_0的拒绝域V应满足 $P(V|H_0\text{真}) \leq \alpha$

即 $V=\{(x_1, x_2, \ldots, x_n):\ U <= -u_\alpha\}$

$u_{0.05}$=1.645 $\qquad U=\dfrac{2960-3000}{80/\sqrt{16}}=-2<-1.645.$

因而拒绝H_0.



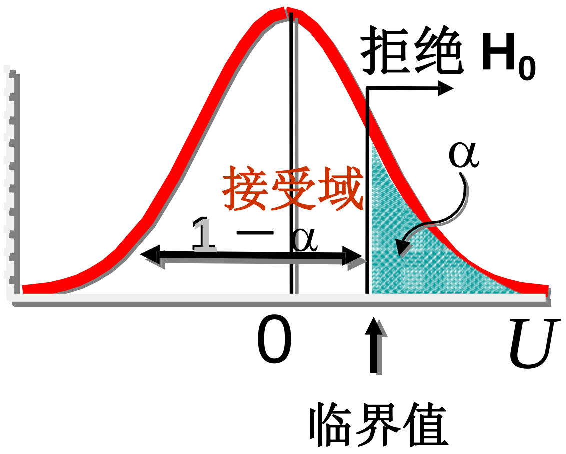
左侧检验

右侧检验：$H_0: \mu \leq 3000$， $H_1: \mu > 3000$

给定α, H_0的拒绝域V应满足

$P(V|H_0\text{真}) \leq \alpha$

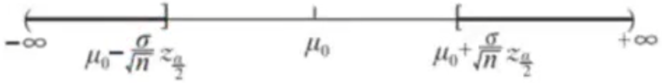
即 $V=\{(x_1, x_2, \ldots, x_n):\ U >= u_\alpha\}$



右侧检验

2. 正态总体下的六大检验及拒绝域

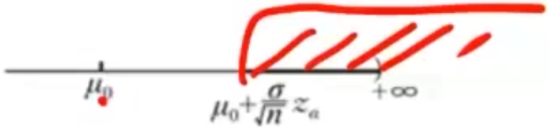
(1)σ^2 已知，μ 未知. $H_0:\mu=\mu_0,H_1:\mu\neq\mu_0$.



(2)σ^2 未知，μ 未知. $H_0:\mu=\mu_0,H_1:\mu\neq\mu_0$.

(3)σ^2 已知，μ 未知. $H_0:\mu\leqslant\mu_0,H_1:\mu>\mu_0$.

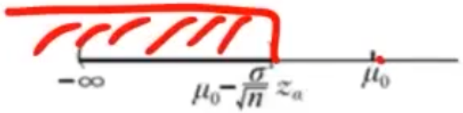
(或写 $\mu=\mu_0$)



(4)σ^2 已知，μ 未知. $H_0:\mu\geqslant\mu_0,H_1:\mu<\mu_0$.

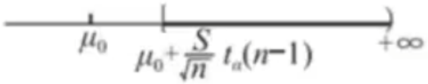
(或写 $\mu=\mu_0$)

单侧的时候是a，而不是a/2



(5)σ^2 未知，μ 未知. $H_0:\mu\leqslant\mu_0,H_1:\mu>\mu_0$.

(或写 $\mu=\mu_0$)



(6)σ^2 未知，μ 未知. $H_0:\mu\geqslant\mu_0,H_1:\mu<\mu_0$.

(或写 $\mu=\mu_0$)

假设检验的 p 值(p –value, probability value)

不同的α有不同的检验结论，为了避免这个问题，引入了 p 值。

假设检验的 p 值是利用样本观测值能够做出拒绝原假设H_0的最小显著性水平。The p-value is the probability of observing data at least as extreme as what has been observed, given that the null hypothesis is true.

p值越小，拒绝原假设的理由越充分。

通常约定: $p \le 0.05$称结果为显著， $p \le 0.01$称结果为高度显著

p 值检验法:

根据样本观测值计算检验的p值，将p值与事先设定的显著性水平α比较大小，作出判断。

(1)如果$p \le \alpha$，则在显著性水平α下**拒绝**原假设H_0

(2)如果$p > \alpha$，则在显著性水平α下**接受**原假设H_0

例 统计软件对检验问题给出检验的 p 值为0.015

α	0.05	0.04	0.03	0.02	0.01
H_0	拒绝	拒绝	拒绝	拒绝	接受

例 统计软件对检验问题给出检验的 p 值为0.03

α	0.05	0.04	0.03	0.02	0.01
H_0	拒绝	拒绝	接受	接受	接受

填空题 1分

设置

若总体 $X \sim N(\mu, 1)$，检验假设 $H_0: \mu = 0, H_1: \mu < 0$，已取得容量为16的样本，$\bar{X}$ 是样本均值，若根据样本观测值，$\bar{x} = -0.5$，则p值为 [填空1] (填写A/B/C/D)

A 0.97725 B 0.95 C 0.05 D 0.02275

已知信息：

- 总体 $X \sim N(\mu, 1)$，方差为1（标准差为1）。
- 检验假设：
  - $H_0: \mu = 0$
  - $H_1: \mu < 0$（左侧检验）
- 样本均值 $\bar{x} = -0.5$，样本容量 $n = 16$。
- 标准误差：$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{16}} = 0.25$。

计算检验统计量 z：

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{-0.5 - 0}{0.25} = -2$$

查找p值：

根据正态分布表：

- $P(Z \leq -2) = 0.02275$。

结论：

p值为 **D. 0.02275**。

标准正态分布表

u	0.00	0.01	0.02	0.03	0.04	0.05
1.9	0.971 28	0.971 93	0.972 57	0.973 20	0.973 81	0.974 41
2.0	0.977 25	0.977 78	0.978 31	0.978 82	0.979 32	0.979 82
2.1	0.982 14	0.982 57	0.983 00	0.983 41	0.983 82	0.984 22

作答

正态总体参数的假设检验

- 单个正态总体
  - 均值的参数检验
  - 方差的参数检验
- 两个正态总体
  - 均值差的参数检验
  - 方差比的参数检验

1 单个正态总体参数的假设检验

设$X_1, X_2, \ldots, X_n$是来自总体$X \sim N(\mu, \sigma^2)$的简单随机样本,

$\overline{X}, S^2$是样本均值和样本方差。

均值的参数检验

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

方差的参数检验

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

（1）方差σ^2已知，关于μ 的检验

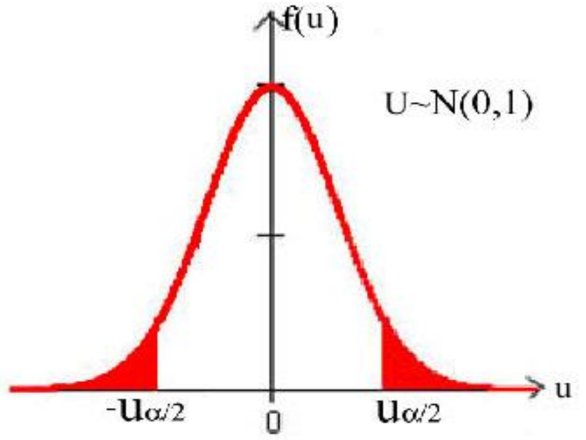
① 原假设与备择假设，$H_0: \mu = \mu_0 \Leftrightarrow H_1:\mu \neq \mu_0$.

② 检验统计量为

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0\text{成立}}{\sim} N(0,1)$$

③ 对给定的显著性水平α，H_0的拒绝域为

$$V = \{(x_1, x_2, \ldots, x_n) : |U| > u_{\alpha/2}\}$$



④ 由样本观察值算出统计量的实测值

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

作出判断：若$|U| > u_{\alpha/2}$，则拒绝H_0，否则接受H_0.

用统计量U来确定拒绝域的检验称为U检验(或Z-test).

U 检验法 (σ^2 已知)

原假设 H_0	备择假设 H_1	检验统计量及其 H_0为真时的分布	拒绝域
$\mu = \mu_0$	$\mu \neq \mu_0$	$U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,\ 1)$	$\lvert U \rvert \geq u_{\frac{\alpha}{2}}$
$\mu \geq \mu_0$	$\mu < \mu_0$		$U \leq -u_{\alpha}$
$\mu \leq \mu_0$	$\mu > \mu_0$		$U \geq u_{\alpha}$

例 化肥厂用自动包装机装化肥，规定每袋标准重量为100，设每袋重量X服从正态分布且标准差$\sigma = 0.9$不变.

某天抽取9袋，测得重量为

99.3， 98.7， 101.2， 100.5， 98.3， 99.7， 102.6， 100.5， 105.1

问机器工作是否正常($\alpha = 0.05$)？

解： H_0：$\mu = 100$； H_1：$\mu \neq 100$

$\boxed{U\text{ 检验法}}$

构造统计量$U = \dfrac{\bar{X}-100}{\sigma/\sqrt{n}} \sim N(0,1)$

$$|\boldsymbol{U}| > \boldsymbol{u}_{\alpha/2}$$

$$U = \frac{100.66-100}{\dfrac{0.9}{\sqrt{9}}} = 2.2 > u_{0.025} = 1.96$$

拒绝

例 某织物强力指标X的均值μ_0=21公斤. 改进工艺后生产一批织物，今从中取30件，测得$\bar{X}$=21.55公斤. 假设强力指标服从正态分布$N(\mu,\ \sigma^2)$, 且已知σ=1.2公斤，问在显著性水平α=0.01下，新生产织物比过去的织物强力是否有提高?

解:提出假设: $H_0: \mu \le 21; \quad H_1: \mu > 21$

取统计量$U = \dfrac{\bar{X}-21}{\sigma/\sqrt{n}}$

拒绝域 V: $\quad U > u_{0.01}$ =2.33

代入 $\boldsymbol{\sigma}$ =1.2, n=30，并由样本值计算得统计量U的实测值 U=2.51>2.33

落入拒绝域

故拒绝原假设H_0.

例 某公司生产的一种灯泡, 其寿命 X 服从正态分布 $N(\mu, 300^2)$. 很长时间以来，灯泡的平均寿命 μ 一直没有超过 2000 h．现在采用新工艺后，从所生产的灯泡中抽取 16 只, 测得平均寿命为 2168 h. 问采用新工艺后, 灯泡的寿命是否有显著提高? ($\alpha = 0.05$)

解 这是正态总体均值的单侧检验问题，需检验假设

$$H_0: \mu \leqslant 2000,\ H_1: \mu > 2000.$$

选取检验统计量

$$U = \frac{\bar{X} - 2000}{300/\sqrt{n}}$$

对显著性水平 $\alpha = 0.05$, 由正态分布表得 $u_{0.05} = 1.645$, 故得拒绝域为

$$V = \{(x_1, \cdots, x_{16}): U > 1.645\}.$$

据样本知 $U = \frac{2168-2000}{300/\sqrt{16}} = 2.24 > 1.645$,

因而拒绝 $H_0: \mu \leqslant 2000$，即可以认为新灯泡的寿命有显著提高.

(2) 方差σ^2未知，关于μ的检验

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

此时不能用U作为检验统计量，因为σ^2未知，U不是统计量，不能根据U确定拒绝域.

选取$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$作为检验统计量.

当H_0成立时，$T \sim t(n-1)$，统计量T的绝对值偏大时，应拒绝H_0，由于

$$P(|T| > t_{\alpha/2}(n-1) \mid H_0\text{真}) = \alpha.$$

因此对给定的显著性水平α，H_0的拒绝域为

$$V = \{(x_1, x_2, \ldots, x_n) : |T| > t_{\alpha/2}(n-1)\}.$$

用统计量T来确定拒绝域的检验称为t检验(T-test).

T 检验法 (σ^2 未知)

原假设 H_0	备择假设 H_1	检验统计量及其 H_0为真时的分布	拒绝域
$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}} \sim t\,(n-1)$	$\lvert T\rvert \geq t_{\frac{\alpha}{2}}(n-1)$
$\mu \geq \mu_0$	$\mu < \mu_0$		$T \leq -t_{\alpha}(n-1)$
$\mu \leq \mu_0$	$\mu > \mu_0$		$T \geq t_{\alpha}(n-1)$

例　金属锰的熔化点 X 服从正态分布，测量5次得熔化点为：1267，　1271，　1256，　1265，　1254

能否认为熔化点为 $1260(\alpha = 0.05)$？

解：　$X \sim N(\mu, \sigma^2)$, $\boldsymbol{\sigma}^2$ **未知.**

$$H_0 : \mu = 1260; \quad H_1 : \mu \neq 1260$$

接受

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

$$t = \frac{\overline{X} - 1260}{S/\sqrt{5}} = 0.796 < t_{0.025}(4) = 2.776$$

(3) μ 已知，关于σ^2的检验

χ^2检验法(μ 已知)

原假设 H_0	备择假设 H_1	检验统计量及其在H_0为真时的分布	拒绝域
$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2}\sim\chi^2(n)$	$\chi^2\leq\chi^2_{1-\frac{\alpha}{2}}(n)$ 或 $\chi^2\geq\chi^2_{\frac{\alpha}{2}}(n)$
$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$		$\chi^2\leq\chi^2_{1-\alpha}(n)$
$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$		$\chi^2\geq\chi^2_{\alpha}(n)$

（4）均值μ未知，关于σ^2的检验

① 原假设及备择假设

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

② 由于样本方差S^2是总体方差的无偏估计，

可选取

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

作为检验统计量.

当H_0成立时，$\chi^2 \sim \chi^2(n-1)$.

χ^2取值较小或较大时，拒绝H_0

③ 由于 $P(\chi^2 < \chi^2_{1-\alpha/2}(n-1) \mid H_0\text{真}) = \frac{\alpha}{2}$,

$$P(\chi^2 > \chi^2_{\alpha/2}(n-1) \mid H_0\text{真}) = \frac{\alpha}{2}.$$

对给定的显著性水平α，拒绝域为

$$V = \left\{\chi^2 < \chi^2_{1-\alpha/2}(n-1) \text{ 或 } \chi^2 > \chi^2_{\alpha/2}(n-1)\right\}$$

④ 由样本算出统计量的值 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$

若$\chi^2 < \chi^2_{1-\alpha/2}(n-1)$ 或 $\chi^2 > \chi^2_{\alpha/2}(n-1)$, 则拒绝$H_0$,否则接受H_0.

用统计量χ^2来确定拒绝域的检验法称为χ^2检验.

χ^2检验法(μ 未知)

原假设 H_0	备择假设 H_1	检验统计量及其在H_0为真时的分布	拒绝域
$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$	$\chi^2\le\chi^2_{1-\frac{\alpha}{2}}(n-1)$ 或 $\chi^2\ge\chi^2_{\frac{\alpha}{2}}(n-1)$
$\sigma^2\ge\sigma_0^2$	$\sigma^2<\sigma_0^2$		$\chi^2\le\chi^2_{1-\alpha}(n-1)$
$\sigma^2\le\sigma_0^2$	$\sigma^2>\sigma_0^2$		$\chi^2\ge\chi^2_{\alpha}(n-1)$

例　已知维尼纶纤度 X 在正常情况下服从正态分布 $N(\mu, 0.048^2)$. 现在测了5 根纤维，其纤度分别为:

1.44， 1.36， 1.40， 1.55， 1.32，

问：产品的精度是否有显著的变化 $(\alpha = 0.05)$?

解: $H_0 : \sigma^2 = 0.048^2; \quad H_1 : \sigma^2 \neq 0.048^2.$

拒绝

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

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{0.03112}{0.048^2} = 13.51 > \chi^2_{0.025}(4) = 11.143$$

$$\chi^2_{0.975}(4) = 0.485$$

例 一种饼干的包装盒上标注净重200g，假设包装盒的重量为定值，且设饼干净重服从$N(\mu,\sigma^2)$，μ, σ^2均未知. 现从货架上取来3盒，称得毛重(单位：g)为 233，215，221，根据这些数据是否可以认为这种包装饼干的标准差超过6g？

解: $H_0: \sigma^2 \le 36,\quad H_1: \sigma^2 > 36,$

拒绝域: $\dfrac{(n-1)S^2}{\sigma_0^2} \ge \chi_\alpha^2(n-1)$

查表得: $\chi_{0.05}^2\left(2\right) = 5.991,$

计算得: $s^2=84, \chi^2 = \dfrac{(3-1)\times 84}{36} = 4.667 < 5.991$

不能拒绝原假设，即认为标准差不超过6g.

小结

单个正态总体$X \sim N(\mu, \sigma^2)$均值μ的检验(表1)

条件	检验问题	检验统计量	H_0真时检验统计量分布	拒绝域
σ^2已知	$H_0:\mu=\mu_0 \leftrightarrow H_1:\mu\neq\mu_0$	$U=\dfrac{X-\mu_0}{\sigma/\sqrt{n}}$	$N(0,1)$	$\lvert U\rvert > u_{\alpha/2}$
	$H_0:\mu\leq\mu_0 \leftrightarrow H_1:\mu>\mu_0$		u检验	$U > u_\alpha$
	$H_0:\mu\geq\mu_0 \leftrightarrow H_1:\mu<\mu_0$			$U < -u_\alpha$
σ^2未知	$H_0:\mu=\mu_0 \leftrightarrow H_1:\mu\neq\mu_0$	$T=\dfrac{X-\mu_0}{S/\sqrt{n}}$	$t(n-1)$	$\lvert T\rvert > t_{\alpha/2}$
	$H_0:\mu\leq\mu_0 \leftrightarrow H_1:\mu>\mu_0$		t 检验	$T > t_\alpha$
	$H_0:\mu\geq\mu_0 \leftrightarrow H_1:\mu<\mu_0$			$T < -t_\alpha$

单个正态总体X：$N(\mu,\sigma^2)$方差σ^2的检验(表2)

条件	检验问题	检验统计量	H_0真时检验统计量分布	拒绝域
μ已知	$H_0:\sigma^2=\sigma_0^2 \leftrightarrow H_1:\sigma^2\neq\sigma_0^2$	$\chi^2=\dfrac{\sum(x_i-\mu_0)^2}{\sigma_0^2}$	$\chi^2(n)$	$\chi^2<\chi^2_{1-\alpha/2}$ 或$\chi^2>\chi^2_{\alpha/2}$
	$H_0:\sigma^2\le\sigma_0^2 \leftrightarrow H_1:\sigma^2>\sigma_0^2$		χ^2检验	$\chi^2>\chi^2_{\alpha}$
	$H_0:\sigma^2\ge\sigma_0^2 \leftrightarrow H_1:\sigma^2<\sigma_0^2$			$\chi^2<\chi^2_{1-\alpha}$
μ未知	$H_0:\sigma^2=\sigma_0^2 \leftrightarrow H_1:\sigma^2\neq\sigma_0^2$	$\chi^2=\dfrac{\sum(x_i-\overline{X})^2}{\sigma_0^2}$	$\chi^2(n-1)$	$\chi^2<\chi^2_{1-\alpha/2}$ 或$\chi^2>\chi^2_{\alpha/2}$
	$H_0:\sigma^2\le\sigma_0^2 \leftrightarrow H_1:\sigma^2>\sigma_0^2$		χ^2检验	$\chi^2>\chi^2_{\alpha}$
	$H_0:\sigma^2\ge\sigma_0^2 \leftrightarrow H_1:\sigma^2<\sigma_0^2$			$\chi^2<\chi^2_{1-\alpha}$

可以看出：对同一个参数的双侧检验、右侧检验和左侧检验, 选用的检验统计量是相同的，但拒绝域是不同的。

- 双侧检验的拒绝域是统计量的取值偏大或偏小
- 右侧检验的拒绝域是统计量的取值偏大
- 左侧检验的拒绝域是统计量的取值偏小.

填空题 1分

设置

一种元件要求其使用寿命不低于1000小时，现在从这批元件中任取25件，测得其寿命平均值为950小时，已知该元件寿命服从标准差为 100小时的正态分布，问这批元件是否合格?

显著性水平 = 0.05.

A 合格 B 不合格

答：[填空1] (填A/B)

α	0.1	0.05	0.04	0.03	0.02	0.01
$u_{\alpha/2}$	1.645	1.96	2.05	2.17	2.32	2.58

这个题目需要找u(a=0.05)，因此是1.645这个

检验统计量公式：

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

代入计算：

$$z = \frac{950 - 1000}{\frac{100}{\sqrt{25}}} = \frac{-50}{20} = -2.5$$

临界值：

查标准正态分布表，显著性水平 $\alpha = 0.05$ 的单侧检验临界值为：

$$z_{0.05} = -1.645$$

比较与决策：

- 计算的 $z = -2.5$ 小于临界值 $z_{0.05} = -1.645$。
- 因此，**拒绝原假设 H_0**。

结论：

在显著性水平 $\alpha = 0.05$ 下，这批元件寿命不符合要求，**答案是 B（不合格）**。

作答

2 两个正态总体参数的假设检验(two-sample hypothesis test)

设 $X_1, X_2, \ldots, X_m$是来自总体$X \sim N(\mu_1, \sigma_1^2)$的简单随机样本;

$Y_1, Y_2, \ldots, Y_n$是来自总体$Y \sim N(\mu_2, \sigma_2^2)$的简单随机样本,

两样本相互独立,$\overline{X}$、$\overline{Y}$为两样本的均值,S_x^2、S_y^2为两样本的方差,

显著性水平为α.

两个正态总体的参数检验,主要有:

① 已知$\boldsymbol{\sigma_1^2}$ 与 $\boldsymbol{\sigma_2^2}$, 检验假设 $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$

② 未知 σ_1^2, σ_2^2, 但 $\sigma_1^2 = \sigma_2^2$, 检验假设 $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$

③ 已知μ_1 与 μ_2, , 检验假设 $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 \neq \sigma_2^2$

④ 未知 μ_1, μ_2, 检验假设 $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 \neq \sigma_2^2$

（1）$\boldsymbol{\sigma_1^2}$ 与 $\boldsymbol{\sigma_2^2}$ 已知, $\boldsymbol{\mu_1-\mu_2}$ 的检验

要检验假设 $\mathrm{H}_0: \mu_1-\mu_2=\delta$, $\mathrm{H}_1: \mu_1-\mu_2 \neq \delta$,

其中 $\delta \geqslant 0$ 是事先给定的常数,通常取 $\delta=0$.

$\mu_1-\mu_2$ 的极大似然估计量是 $\bar{X}-\bar{Y}$, 且 $\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2; \frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)$.

当 H_0 成立时, $\mu_1-\mu_2=\delta$, 检验统计量取为

$$U=\frac{(\bar{X}-\bar{Y})-\delta}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \sim N(0;1)$$

应用 u 检验,对于给定的显著性水平 α, 检验的拒绝域为

$$V=\left\{(x_1,x_2,\cdots,x_m;y_1,y_2,\cdots,y_n): \frac{|(\bar{x}-\bar{y})-\delta|}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}>u_{a/2}\right\}.$$

(2) $\sigma_1^2=\sigma_2^2=\sigma^2$ 但未知, $\mu_1-\mu_2$ 的检验

原假设与备择假设为 $H_0: \mu_1=\mu_2 \Leftrightarrow H_1: \mu_1\neq\mu_2$

统计量$\bar{X}-\bar{Y}$可作为$\mu_1-\mu_2$的估计，

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{1/m+1/n}}\sim t(m+n-2)$$

$$S_W^2=\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}$$

当H_0成立时，可选取检验统计量$T=\dfrac{(\bar{X}-\bar{Y})}{S_W\sqrt{1/m+1/n}}$

$|T|$的值一般不能很大，因此H_0的拒绝域为

$V=\{(x_1,x_2,\ldots,x_n):|T|>t_{\alpha/2}(m+n-2)\}$（$t$检验法中的双侧检验）

例 某种物品在处理前与处理后分别抽取样本分析其含脂率如下:

处理前: 0.19 0.18 0.21 0.30 0.41 0.12 0.27

处理后: 0.15 0.13 0.07 0.24 0.19 0.06 0.08 0.12

假定处理前后的含脂率都服从正态分布, 且方差不变, 问在显著性水平 $\alpha = 0.05$ 条件下，处理前后含脂率的均值是否有显著性差异?

解 由样本计算得 $\bar{x} = 0.24,\ s_1^2 \approx 0.0091,\ m = 7;\ \bar{y} = 0.13,\ s_2^2 \approx 0.0039,\ n = 8.$

$$\sigma_1^2 = \sigma_2^2$$

要检验的假设为 $H_0: \mu_1 = \mu_2,\ H_1: \mu_1 \neq \mu_2.$

选择检验统计量

$$T = \frac{\bar{X}-\bar{Y}}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \overset{H_0\text{成立}}{\sim} t(m+n-2), \text{其中} \quad S_w^2 = \frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}.$$

代入数据计算得 $T \approx 2.68$, 查附表得 $t_{0.025}(13) = 2.1604$, 故 $|T| > t_{0.025}(13)$,

所以拒绝 H_0, 即认为处理前后含脂率的均值有显著性差异.

(3) $\boldsymbol{\sigma_1^2}$与$\boldsymbol{\sigma_2^2}$未知, 成对数据$\boldsymbol{\mu_1-\mu_2}$的检验

例 在机器学习中，对两个学习器 A 和 B , 使用 k 折交叉验证法得到的测试错误率分别为 $\epsilon_1^A,\epsilon_2^A,\dots,\epsilon_k^A$ 和 $\epsilon_1^B,\epsilon_2^B,\dots,\epsilon_k^B$,其中 ϵ_i^A 和 ϵ_i^B 是在相同的第 i 折测试集上得到的结果。假定测试错误率服从正态分布, 比较检验两个学习器的性能，即考察如下检验问题:

$$H_0: \epsilon_i^A=\epsilon_i^B \ \text{ vs } H_1: \epsilon_i^A\neq\epsilon_i^B$$

对 k 折交叉验证产生的 k 对测试错误率的每对结果求差, $d_i=\epsilon_i^A-\epsilon_i^B$; 若两个学习器性能相同,则它们使用相同的测试集得到的测试错误率应相同, 差值均值应为零. 因此,可根据差值 $d_1,d_2,\dots,d_k$ 来对 "学习器 A 与 B 性能相同"这个假设做成对 t 检验(paired t-tests), 计算出差值的均值 $\bar{d}$ 和方差 ${s_d}^2$,

$$\bar{d}=\frac{1}{k}\sum_{i=1}^{k} d_i,\ s_d=\left(\frac{1}{k-1}\sum_{i=1}^{k}\left(d_i-\bar{d}\right)^2\right)^{1/2}$$

如果H_0成立，则检验统计量 $t=\bar{d}/\left(s_d/\sqrt{k}\right)\sim t(k-1)$

在给定显著性水平 α 下,该检验问题的拒绝域是

$$W=\{|t|\geqslant t_{\alpha/2}(k-1)\},$$

若统计量 t 的值小于临界值 $t_{\alpha/2,k-1}$,则假设不能被拒绝,即认为两个学习器的性能没有显著差别; 否则可认为两个学习器的性能有显著差别, 且平均错误率较小的那个学习器性能较优.

(4) $\boldsymbol{\mu_1}$ 与 $\boldsymbol{\mu_2}$ 已知, $\boldsymbol{\sigma_1^2}$ 与 $\boldsymbol{\sigma_2^2}$ 差异的检验

构造统计量

$$F=\frac{\sum_{i=1}^{m}(X_i-\mu_1)^2/\sigma_1^2/m}{\sum_{j=1}^{n}\left(Y_j-\mu_2\right)^2/\sigma_2^2/n}=\frac{\sigma_2^2}{\sigma_1^2}\frac{n\sum_{i=1}^{m}(X_i-\mu_1)^2}{m\sum_{j=1}^{n}\left(Y_j-\mu_2\right)^2}\sim F(m,n),$$

相应的检验统计量 $F_0=\dfrac{n\sum_{i=1}^{m}(X_i-\mu_1)^2}{m\sum_{j=1}^{n}\left(Y_j-\mu_2\right)^2}$.

- 对于**双侧检验问题**

$$H_0:\sigma_1^2=\sigma_2^2\leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2,$$

当 H_0 成立时, $F_0\sim F(m,n)$, 拒绝域可取为

$$V=\left\{(x_1,\cdots,x_m;y_1,\cdots,y_n):F_0<F_{1-\frac{\alpha}{2}}(m,n)\text{ 或 }F_0>F_{\frac{\alpha}{2}}(m,n)\right\},$$

其中 $F_{1-\frac{\alpha}{2}}(m,n),F_{\frac{\alpha}{2}}(m,n)$ 分别为水平 $1-\frac{\alpha}{2}$ 与 $\frac{\alpha}{2}$ 的 $F(m,n)$ 分布的上侧分位数;

- 对于**右侧检验问题**

$$H_0: \sigma_1^2 \leqslant \sigma_2^2 \leftrightarrow H_1: \sigma_1^2 > \sigma_2^2,$$

拒绝域为 $V = \{(x_1, \cdots, x_m, y_1, \cdots, y_n): F_0 > F_\alpha(m, n)\},$

其中 $F_\alpha(m,n)$ 为水平 α 的 $F(m,n)$ 分布的上侧分位数;

- 对于**左侧检验问题**

$$H_0: \sigma_1^2 \geqslant \sigma_2^2 \leftrightarrow H_1: \sigma_1^2 < \sigma_2^2,$$

拒绝域为 $V = \{(x_1, \cdots, x_m, y_1, \cdots, y_n): F_0 < F_{1-\alpha}(m, n)\},$

其中 $F_{1-\alpha}(m,n)$ 为水平 $1-\alpha$ 的 $F(m,n)$ 分布的上侧分位数.

(5) $\boldsymbol{\mu_1}$ 与 $\boldsymbol{\mu_2}$ 均末知, $\boldsymbol{\sigma_1^2}$ 与 $\boldsymbol{\sigma_2^2}$ 差异的检验

这里均值μ_1与μ_2未知，检验假设 H_0: $\sigma_1^2=\sigma_2^2$; H_1: $\sigma_1^2\neq\sigma_2^2$

由于样本方差 S^2 是总体方差 σ^2 的良好估计, 一般情况下二者相差不大.

当 H_0 成立时, 比值 $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ 的值不应过大, 也不应过小, 否则就应拒绝 H_0.

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

当 H_0 成立时,检验统计量$F=\frac{S_1^2}{S_2^2}\sim F(m-1,n-1)$.

若 F 的取值偏大或偏小, 应拒绝 H_0. 对于给定的显著性水平 α, 拒绝域为

$$V=\left\{(x_1,x_2,\cdots,x_n):F>F_{\alpha/2}(m-1,n-1)\text{ 或 }F<F_{1-\alpha/2}(m-1,n-1)\right\}.$$

例 从某机床厂的两台机器所加工的同一种零件中，分别抽取11个和9个样本进行测量，测得其尺寸（单位:cm）如下：

第一台机器：6.2, 5.7, 6.0, 6.3, 6.5, 6.0,
5.7, 5.8, 6.0, 5.8, 6.0;

第二台机器：5.6, 5.7, 5.9, 5.5, 5.6, 6.0,
5.8, 5.5, 5.7.

已知零件尺寸服从正态分布，问在显著性水平α=0.05 的条件下，加工精度（方差）是否有显著差异？

解　由题设可得　$\bar{X}=6.0$，$S_1^2=0.064$，$m=11$;

$\bar{Y}=5.7$，$S_2^2=0.030$，$n=9$.

检验假设　$H_0: \sigma_1^2=\sigma_2^2 \Leftrightarrow H_1: \sigma_1^2 \neq \sigma_2^2$

选取检验统计量

$$F=S_1^2/S_2^2 \overset{H_0\text{成立}}{\sim} F(m-1,n-1)$$

当$\alpha=0.05$时，查表得

$$F_{0.025}(10,8)=4.3,\ F_{0.975}(10,8)=1/F_{0.025}(8,10)=0.26$$

H_0的拒绝域为：$V=\{F<0.26\text{或}F>4.3\}$

$$F=S_1^2/S_2^2=0.064/0.03=2.13$$

$$0.26<F<4.3$$

因而在显著性水平$\alpha=0.05$的条件下，接受原假设H_0，即认为两机器的精度无显著性差异.

例 甲、乙两台机床生产同一型号的滚珠. 今从它们生产的滚珠中分别抽取了 8 个与 9 个,测得直径(单位: mm)的数据如下，假定滚珠直径服从正态分布，问两台机床生产的滚珠直径是否可以认为具有同一分布 $(\alpha = 0.05)$?

甲机床	15.0	14.5	15.2	15.5	14.8	15.1	15.2	14.8	
乙机床	15.2	15.0	14.8	15.2	15.0	15.0	15.1	14.8	14.8

解 如果是服从同一分布,那么 $\sigma_1^2 = \sigma_2^2$ 及 $\mu_1 = \mu_2$,因此要先检验方差齐性,再检验均值是否相等.

(1) 由所给数据算得 $\bar{x}=15.0125, \bar{y}=14.9889, s_1^2=0.0955, s_2^2=0.0261$, 且 $m=8, n=9$, $s_w^2=0.0585$, 在显著性水平 $\alpha=0.05$ 下,检验方差齐性假设

$$\mathrm{H}_0: \sigma_1^2=\sigma_2^2,\ \mathrm{H}_1: \sigma_1^2 \neq \sigma_2^2.$$

应用 F 检验, 得到检验统计量的观测值

$$\frac{s_1^2}{s_2^2}=\frac{0.0955}{0.0261}=3.6590,$$

而临界值 $F_{0.025}(7,8)=4.53, F_{0.975}(7,8)=\frac{1}{F_{0.025}(8,7)}=\frac{1}{4.90}=0.2041$.

0.2041$<3.6590<4.53$, 因此接受H_0, 即认为它们的方差相等.

(2) 再检验假设

$$H_0: \mu_1 = \mu_2,\ H_1: \mu_1 \neq \mu_2.$$

应用 t 检验, 得到检验统计量的观测值

$$\frac{\bar{x}-\bar{y}}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}} = \frac{15.0125-14.9889}{\sqrt{0.0585}\times\sqrt{\frac{1}{8}+\frac{1}{9}}} = 0.2008.$$

在显著性水平 $\alpha = 0.05$ 下,临界值 $t_{0.025}(15) = 2.1315$,显然有 $|0.2008| < 2.1315$, 因此 不能拒绝 H_0, 即认为均值没有差异.

综合两个结果,可以认为两台机床生产的滚珠直径具有同一分布.

两个正态总体均值的检验

条件	检验问题	检验统计量	H_0真时检验统计量分布	拒绝域
σ_1^2 σ_2^2 已知	$H_0:\mu_1=\mu_2 \leftrightarrow$ $H_1:\mu_1\neq\mu_2$	$U=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}}$	$N(0,1)$	$\lvert U\rvert>u_{\alpha/2}$
	$H_0:\mu_1\le\mu_2 \leftrightarrow$ $H_1:\mu_1>\mu_2$		u检验	$U>u_\alpha$
	$H_0:\mu_1\ge\mu_2 \leftrightarrow$ $H_1:\mu_1<\mu_2$			$U<-u_\alpha$
$\sigma_1^2=\sigma_2^2$ 未知	$H_0:\mu_1=\mu_2 \leftrightarrow$ $H_1:\mu_1\neq\mu_2$	$T=\dfrac{\bar{X}-\bar{Y}}{S_W\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$	$t(m+n-2)$	$\lvert T\rvert>t_{\alpha/2}$
	$H_0:\mu_1\le\mu_2 \leftrightarrow$ $H_1:\mu_1>\mu_2$		t 检验	$T>t_\alpha$
	$H_0:\mu_1\ge\mu_2 \leftrightarrow$ $H_1:\mu_1<\mu_2$			$T<-t_\alpha$

两个正态总体方差的检验

条件	检验问题	检验统计量	H_0真时检验统计量分布	拒绝域
μ_1 μ_2 已知	$H_0:\sigma_1^2=\sigma_2^2 \leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$	$F=\dfrac{\sum(x_i-\mu_1)^2/m}{\sum(y_i-\mu_2)^2/n}$	$F(m,n)$	$F<F_{1-\alpha/2}$ 或$F>F_{\alpha/2}$
	$H_0:\sigma_1^2\leq\sigma_2^2 \leftrightarrow H_1:\sigma_1^2>\sigma_2^2$		F检验	$F>F_\alpha$
	$H_0:\sigma_1^2\geq\sigma_2^2 \leftrightarrow H_1:\sigma_1^2<\sigma_2^2$			$F<F_{1-\alpha}$
μ_1 μ_2 未知	$H_0:\sigma_1^2=\sigma_2^2 \leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$	$F=S_x^2/S_y^2$	$F(m-1,n-1)$	$F<F_{1-\alpha/2}$ 或$F>F_{\alpha/2}$
	$H_0:\sigma_1^2\leq\sigma_2^2 \leftrightarrow H_1:\sigma_1^2>\sigma_2^2$		F检验	$F>F_\alpha$
	$H_0:\sigma_1^2\geq\sigma_2^2 \leftrightarrow H_1:\sigma_1^2<\sigma_2^2$			$F<F_{1-\alpha}$

非正态总体大样本参数检验

- 在正态总体参数检验中，由于检验统计量都有精确分布，因而对样本大小没有限制.
- 对于非正态总体，一般难以找到具有精确分布的检验统计量.
  - 当样本容量 n 较大时，往往用检验统计量的渐近分布代替它的精确分布，从而得到近似的拒绝域. 这里，点估计的渐近正态性起着重要的作用.

对于非正态总体, 要求出统计量的分布是比较困难的, 即使有时理论上可以求出精确的分布, 但其形式复杂而难以应用, 这时一般利用中心极限定理推出 $n \to +\infty$ 下的极限分布. 当样本容量很大时, 应用有关**统计量的极限分布**作为其近似分布.

定理 设总体X的一阶矩、二阶矩存在: $E(X) = \mu, 0 < D(X) = \sigma^2 < +\infty$, $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的样本, 则当 n 充分大时,

(1) 样本均值 $\bar{X}_n$ 近似地服从 $N\left(\mu; \frac{\sigma^2}{n}\right)$,

(2) 样本方差S^2依概率收敛于 σ^2,

(3) $\sqrt{n}\frac{\bar{X}_n-\mu}{S}$ 近似地服从 $N(0;1)$,

其中 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

例 某县早稻收割面积为100万亩（1亩$=666.7\ \mathrm{m}^2$），现随机抽取150亩，得到平均亩产量 $\bar{x}=320\mathrm{kg}$，且样本标准差 $s=100\mathrm{kg}$，问在显著性水平 $\alpha=0.05$ 下，能否预计这100万亩早稻的平均亩产量为 $340\mathrm{kg}$？

解 要检验假设 $\mathrm{H}_0\colon E(X)=340$，$\mathrm{H}_1\colon E(X)\neq 340$，其中 X 表示亩产量．由于 $n=150$ 较大，所以近似地有 $\sqrt{n}\frac{\bar{X}-E(X)}{S}\sim N(0;1)$.

当 H_0 成立时，$E(X)=340$．这时 $U=\sqrt{n}\frac{\bar{X}-340}{S}\sim N(0;1)$.

由 $P\{|U|>u_{\alpha/2}\}\approx\alpha$ 得到检验的拒绝域为

$$V=\left\{(x_1,x_2,\cdots,x_n)\colon\sqrt{150}\,\frac{|\bar{x}-340|}{s}>u_{\alpha/2}\right\}.$$

$\bar{x}=320, s=100$，检验统计量的观测值是

$$\sqrt{150}\,\frac{|320-340|}{100}=2.4495,$$

由 $\alpha=0.05$，查表可得上分位数 $u_{0.025}=1.96$．显然 $2.4495>1.96$，因而拒绝 H_0，即不能预计这 100 万亩的平均亩产量为 $340\mathrm{kg}$.

例 某厂生产一批产品, 质量检查规定, 其次品率不大于 1% 才能出厂销售. 现从这批产品中随机地抽取 100 件进行检查, 发现有 3 件次品, 问这批产品能否出厂? ($\alpha = 0.05$)

解 考虑厂方利益, 要检验假设

$$\mathrm{H}_0: p \leqslant 0.01, \mathrm{H}_1: p > 0.01,$$

其中 p 为该批产品的次品率. 总体 X 服从 0-1 分布 $B(1,p)$, p 的极大似然估计量是 $\bar{X}$, 近似地有

$$\sqrt{100}\frac{\bar{X}-p}{\sqrt{p(1-p)}} \sim N(0;1).$$

当 H_0 成立时 $p \leqslant 0.01$, 取检验统计量为 $U = \sqrt{100}\frac{\bar{X}-0.01}{\sqrt{0.01\times(1-0.01)}}$.

在显著性水平 α 下, 检验的拒绝域为

$$V \approx \left\{(x_1, x_2, \cdots, x_n): \sqrt{100}\frac{\bar{x}-0.01}{\sqrt{0.01\times(1-0.01)}} > u_\alpha\right\}.$$

$\bar{x} = 0.03, \alpha = 0.05$, 检验统计量的观测值为 $\sqrt{100} \times \frac{0.03-0.01}{\sqrt{0.0099}} = 2.01$, 而 $u_\alpha = u_{0.05} = 1.645$. 显然, $2.01 > 1.645$, 因而拒绝 H_0, 即这批产品不能出厂销售.

例 在两种工艺条件下纺出一批细纱，现随机地各抽取100个样品测试其强力(单位:牛)，经计算得到在这两种工艺条件下样本均值和样本标准差为：甲工艺: $\bar{x} = 2.80, s_1 = 0.280$，乙工艺: $\bar{y} = 2.86, s_2 = 0.285$，

问在 $\alpha = 0.05$ 下, 两种工艺条件纺出的细纱的平均强力有无显著差异?

解 这是检验两个总体均值是否相等的问题, 即检验假设

$$\mathrm{H}_0: E(X) = E(Y), \quad \mathrm{H}_1: E(X) \neq E(Y),$$

其中 X、Y 分别表示甲、乙两种工艺条件纺出的细纱总体.

取检验统计量 $U = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}}$, 当 H_0 成立时, U 的渐近分布为 $N(0;1)$.

在显著性水平 α 下, 检验的拒绝域为 $V \approx \left\{(x_1, x_2, \cdots, x_m; y_1, y_2, \cdots, y_n): \frac{|\bar{x}-\bar{y}|}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}} > u_{\alpha/2}\right\}$.

$m = n = 100, \bar{x} = 2.80, \bar{y} = 2.86, s_1^2 = (0.280)^2, s_2^2 = (0.285)^2$，由此得到检验统计量的观测值为 $\frac{|2.80-2.86|}{\sqrt{\frac{(0.280)^2+(0.285)^2}{100}}} = 1.5018$，临界值 $u_{0.025} = 1.96$. 显然, $1.5018 < 1.96$, 因而不能拒绝 H_0, 即在这两种工艺条件下纺出的细纱的平均强力无显著差异.

例 为确定某种除虫剂对植物的影响，取1000株植物做试验. 在没有施这种除虫剂的 100 株植物中，有53株长势良好；在施除虫剂的900株中，有689株长势良好. 问在显著性水平 $\alpha = 0.01$ 下这种除虫剂的效果是否显著.

解 根据题意要检验假设 $\mathrm{H}_0: p_1 \geqslant p_2$, $\mathrm{H}_1: p_1 < p_2$,

其中 p_1 是没有施除虫剂的植物中长势良好的比例， p_2 是施除虫剂的植物中长势良好的比例. 取检验统计量

$$T = \frac{\bar{X}_m - \bar{Y}_n}{\sqrt{\frac{\bar{X}_m(1-\bar{X}_m)}{m} + \frac{\bar{Y}_n(1-\bar{Y}_n)}{n}}},$$

当 H_0 成立时，T的分布近似为$N(0;1)$. 由所给数据知，$m = 100, n = 900, \bar{x}_m = 0.530, \bar{y}_n = 0.766$, 从而得到检验统计量的观测值

$$\frac{\bar{x}_m - \bar{y}_n}{\sqrt{\frac{\bar{x}_m(1-\bar{x}_m)}{m} + \frac{\bar{y}_n(1-\bar{y}_n)}{n}}} = \frac{0.530 - 0.766}{\sqrt{\frac{0.53 \times 0.47}{100} + \frac{0.766 \times 0.234}{900}}} = -4.550,$$

临界值 $-u_\alpha = -u_{0.01} = -2.326$. 显然, $-4.550 < -2.326$, 因此拒绝 H_0, 即这种除虫剂的效果是显著的.

检验的优劣

假设检验的两类错误

由于样本的随机性, 在使用一个检验时, 作出“接受 H_0”或“拒绝 H_0”的判断并不证明假设 H_0一定正确或一定不正确, 而只是表明检验使用者对假设 H_0的一种倾向性意见. 使用一个检验会面临犯两种错误的可能性:

- 当 H_0实际上成立, 但由于样本观测值落在拒绝域W_1内而拒绝 H_0, 这种“弃真错误”称为**第一类错误**;
- 当 H_0实际上不成立, 但由于样本观测值落在拒绝域W_1外而接受 H_0. 这种“取伪错误”称为**第二类错误**.

判断结果	实际情况	
	H_0 成立	H_1 成立
接受 H_0	正确$(1-\alpha)$	第二类错误$(1-\beta)$
拒绝 H_0	第一类错误(α)	正确(β)

- 犯第一类错误的概率为

$$P\{拒绝\ \mathrm{H}_0 \mid \mathrm{H}_0成立\}$$
$$= P\{(X_1, X_2, \cdots, X_n) \in W_1 \mid \mathrm{H}_0成立\}$$

称为**检验的 I 类风险**;

- 犯第二类错误的概率为

$$P\{接受\mathrm{H}_0 \mid \mathrm{H}_0不成立\}$$
$$= P\{(X_1, X_2, \cdots, X_n) \notin W_1 \mid \mathrm{H}_0不成立\}$$

称为**检验的 II 类风险**.

机器学习中的混淆矩阵(Confusion matrix)

Predicted Value (predicted by the test)		Actual Value (as confirmed by experiment)	
		positives	negatives
	positives	**TP** True Positive	**FP** False Positive
	negatives	**FN** False Negative	**TN** True Negative

问题：

设总体 $X \sim N(\mu;\sigma^2)$, 其中 μ 与 σ^2 均未知, 在显著性水平 α 下检验假设

$$\mathrm{H}_0: \mu = \mu_0,\ \mathrm{H}_1: \mu \neq \mu_0,$$

其中 μ_0 是常数, 取检验统计量为 $T = \sqrt{n}\frac{\bar{X}-\mu_0}{S}$,

- 拒绝域取为$W_1 = \left\{(x_1, x_2, \cdots, x_n) \mid \sqrt{n}\frac{|\bar{x}-\mu_0|}{s} > t_{\alpha/2}(n-1)\right\}$
- 由于 $P\{T < t_{1-2\alpha/3}(n-1)\text{ 或 } T > t_{\alpha/3}(n-1)\} = \alpha$, 拒绝域也可以取为

$$W_2 = \left\{(x_1, x_2, \cdots, x_n) \mid \sqrt{n}\frac{\bar{x}-\mu_0}{s} < t_{1-2\alpha/3}(n-1)\ \text{ 或 }\ \sqrt{n}\frac{\bar{x}-\mu_0}{s} > t_{\alpha/3}(n-1)\right\}.$$

为什么检验的拒绝域通常取为 W_1 呢?

在统计学中， 拒绝域的选取准则：在保证I类风险至多为 α 的前提下， 要使 II 类风险尽可能小(功效尽可能大).

功效函数

设 W_1 是检验 ϕ $\mathrm{H}_0: \theta \in \Theta_0, \mathrm{H}_1: \theta \in \Theta_1$的拒绝域，函数

$$\beta(\theta) = P_\theta\{(X_1, X_2, \cdots, X_n) \in W_1\}, \theta \in \Theta$$

称为检验 ϕ 的**功效函数(power function)**, 有时记为 $\beta_\phi(\theta)$.其中 P_θ 表示按总体参数取 θ 值时的分布来计算随机事件的概率。

- 当 $\theta \in \Theta_0$时, $\beta_\phi(\theta)$ 是检验 ϕ 的 I 类风险;
- 当 $\theta \in \Theta_1$时, $\beta_\phi(\theta)$ 是检验 ϕ 对于备选假设 H_1 在 θ 处的**功效**, $1 - \beta_\phi(\theta)$ 是检验 ϕ 的 II 类风险.
- 功效函数通过拒绝域 W_1 把两类风险包含在一个统一的形式之中.

填空题 2分

设置

当 $\theta \in \Theta_0$ 时, 检验 ϕ 的功效函数 $\beta(\theta)$ 增大时, 检验 ϕ 的 I 类风险会变 [填空1] （大/小）

当 $\theta \in \Theta_1$ 时, 检验 ϕ 的功效函数 $\beta(\theta)$ 增大时, 检验 ϕ 的 II 类风险会变 [填空2] （大/小）

作答

例 设 $(X_1, X_2, \cdots, X_n)$ 是取自正态总体 $N(\mu;1)$ 的样本,其中 μ 末知,但 $\mu \in \Theta = \{\mu_0, \mu_1\}$, 且 $\mu_0 < \mu_1$. 对于检验假设 $\mathrm{H}_0\colon \mu = \mu_0$, $\mathrm{H}_1\colon \mu = \mu_1$, H_0 与 H_1 都是简单假设,应用 u 检验得到拒绝域为 $W_1 = \left\{(x_1, x_2, \cdots, x_n) \mid \sqrt{n}(\bar{x} - \mu_0) > u_a\right\}$, 其中 α 为显著性水平.

这个检验的功效函数为 $\beta(\mu) = P_\mu\{(X_1, X_2, \cdots, X_n) \in W_1\} = P_\mu\left\{\sqrt{n}(\bar{X} - \mu_0) > u_\alpha\right\}$

$$= P_\mu\left\{\sqrt{n}(\bar{X} - \mu) > u_a - \sqrt{n}(\mu - \mu_0)\right\} = 1 - P_\mu\left\{\sqrt{n}(\bar{X} - \mu) \leqslant u_\alpha - \sqrt{n}(\mu - \mu_0)\right\}$$

由于 $\bar{X} \sim N\left(\mu; \frac{1}{n}\right)$, 故 $\sqrt{n}(\bar{X} - \mu) \sim N(0;1)$, 记 $\Phi(x)$ 为标准正态分布 $N(0;1)$ 的分布函数, 则 $\beta(\mu) = 1 - \Phi\left(u_a - \sqrt{n}(\mu - \mu_0)\right), \mu \in \Theta = \{\mu_0; \mu_1\}$.

$$1 - \beta(\mu_1) = \Phi\left(u_a - \sqrt{n}(\mu_1 - \mu_0)\right)$$

- 对于固定的样本容量 n, 当 α 减小(即 I 类风险减小) 时, u_α 增大, 使 $1 - \beta(\mu_1)$ 增大, 即II 类风险增大; 反之, 若 α 增大, u_α 减小, 使 $1 - \beta(\mu_1)$ 减小,即II类风险减小. 即当样本容量固定时, I类风险变小会使 II 类风险增大,而II类风险变小会使 I类风险变大.
- 在给定显著性水平 α 下,要使 II 类风险不大于 ε, 则只有确定适当的样本容量 n, 使下式成立:

$$\Phi\left(u_\alpha - \sqrt{n}(\mu_1 - \mu_0)\right) \leqslant \varepsilon, \text{即 } n \geqslant \left(\frac{u_\alpha - u_{1-\varepsilon}}{\mu_1 - \mu_0}\right)^2$$

例 设 $(X_1, X_2, \cdots, X_n)$ 是取自正态总体 $N(\mu;1)$ 的样本,其中 μ 末知, $\mu \in (-\infty, +\infty)$. 要检验假设 $\mathrm{H}_0: \mu = \mu_0$, $\mathrm{H}_1: \mu \neq \mu_0$, 应用 u 检验得到拒绝域为

$$W_1 = \left\{(x_1, x_2, \cdots, x_n) | \sqrt{n} |\bar{x} - \mu_0 | > u_{\alpha/2}\right\}.$$

这个 u 检验的功效函数为

$$\beta(\mu) = P_\mu\{(X_1, X_2, \cdots, X_n) \in W_1\} = P_\mu\left\{\sqrt{n}\left|\bar{X} - \mu_0\right| > u_{\alpha/2}\right\} = 1 - P_\mu\left\{\sqrt{n}\left|\bar{X} - \mu_0\right| \leqslant u_{\alpha/2}\right\}$$

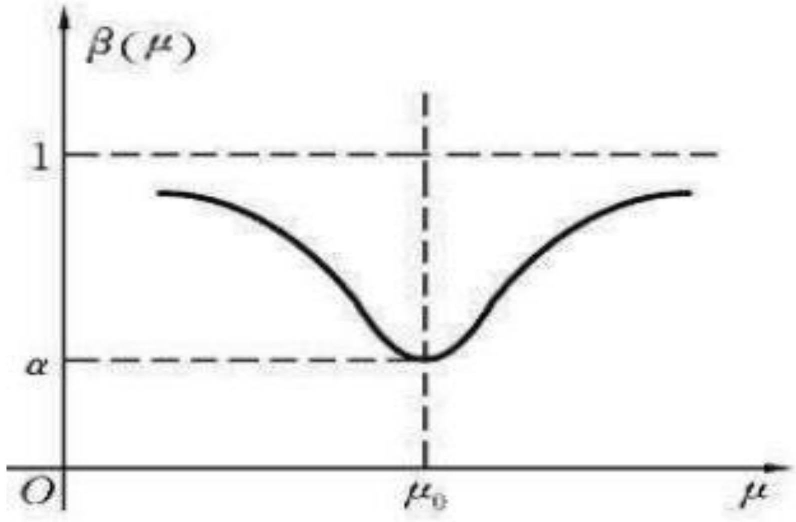
$$= 1 - P_\mu\left\{-u_{\alpha/2} - \sqrt{n}(\mu - \mu_0) \leqslant \sqrt{n}(\bar{X} - \mu) \leqslant u_{\alpha/2} - \sqrt{n}(\mu - \mu_0)\right\}$$

$$= 1 - \left[\Phi\left(u_{\alpha/2} - \sqrt{n}(\mu - \mu_0)\right) - \Phi\left(-u_{\alpha/2} - \sqrt{n}(\mu - \mu_0)\right)\right]$$

$$= \Phi\left(-u_{\alpha/2} + \sqrt{n}(\mu - \mu_0)\right) + \Phi\left(-u_{\alpha/2} - \sqrt{n}(\mu - \mu_0)\right), \quad -\infty < \mu < +\infty$$

$$\beta(\mu_0) = \Phi\left(-u_{a/2}\right) + \Phi\left(-u_{a/2}\right) = \alpha$$

当 $\mu \neq \mu_0$ 时, $\beta(\mu) > \alpha$



这个 u 检验的 II 类风险为

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

- 对于固定的样本容量 n，当 $\mu\to\pm\infty$ 时, $1-\beta(\mu)\to 0$，故当实际的总体均值μ 偏离原假设μ_0 较大时, 这个u 检验的效果较好。

- 当实际的总体均值 μ 接近原假设μ_0时,由于 $\lim\limits_{\mu\to\mu_0}(1-\beta(\mu))=1-\alpha$， α 较小, $1-\alpha$ 较大,这个 u 检验的效果不好.

这表明,无论样本容量 n 多大, 要想对所有的 $\mu\in H_1$, II 类风险都很小是不可能的.

- 可以确定样本容量 n, 使得当 $|\mu-\mu_0|\geqslant\delta(\delta>0)$ 时, II 类风险不大于 ε, n 应满足下式:

$$\Phi\left(u_{\alpha/2}-\sqrt{n}\delta\right)-\Phi\left(-u_{\alpha/2}-\sqrt{n}\delta\right)\leqslant\varepsilon$$

通常, 因 n 较大, 故有 $\Phi\left(-u_{\alpha/2}-\sqrt{n}\delta\right)\approx 0$, 从而近似地有 $\Phi\left(u_{\alpha/2}-\sqrt{n}\delta\right)\leqslant\varepsilon$.

只要样本容量满足 $u_{\alpha/2}-\sqrt{n}\delta\leqslant u_{1-\varepsilon}$, 即 $n\geqslant\left(\frac{u_{\alpha/2}-u_{1-\varepsilon}}{\delta}\right)^2$， 就能使 $\mu\in H_1$ 且 $|\mu-\mu_0|\geqslant\delta$ 时,犯第二类错误的概率不超过 ε.

例 设 $(X_1, X_2, \cdots, X_n)$ 是取自正态总体$N(\mu; 1)$的样本,其中μ未知,但 $\mu \in (-\infty, +\infty)$. 要检验复合假设

$$H_0: \mu \leqslant \mu_0,\ H_1: \mu > \mu_0.$$

应用 u 检验得到拒绝域为

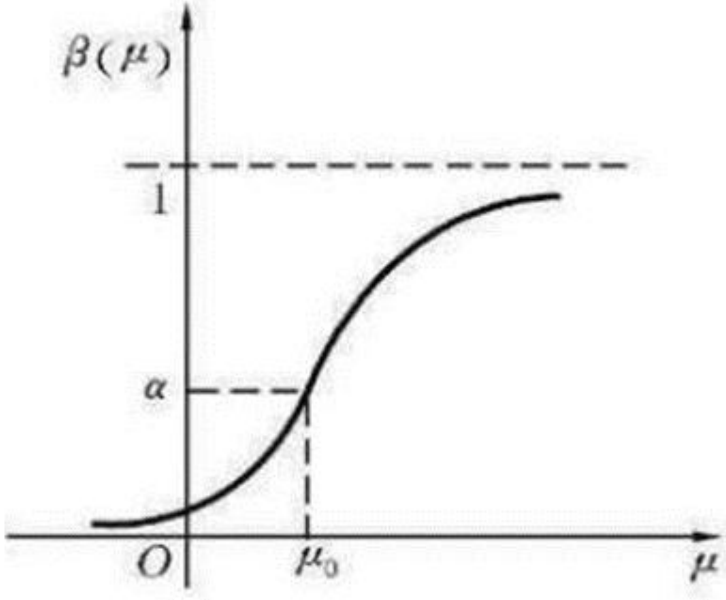
$$W_1 = \{(x_1, x_2, \cdots, x_n) \mid \sqrt{n}(\bar{x} - \mu_0) > u_\alpha\}.$$

这个 u 检验的功效函数为

$$\beta(\mu) = P_\mu\{\sqrt{n}(\bar{X} - \mu_0) > u_\alpha\}$$

$$= P_\mu\{\sqrt{n}(\bar{X} - \mu) > u_\alpha - \sqrt{n}(\mu - \mu_0)\}$$

$$= 1 - \Phi\left(u_\alpha - \sqrt{n}(\mu - \mu_0)\right), \mu \in (-\infty, +\infty)$$



- 当 H_0 成立, 即 $\mu \leqslant \mu_0$ 时, $\beta(\mu) \leqslant \alpha$, 这个 u 检验符合显著性水平为 α 的要求, 即 I类风险至多为 α.

- 这个 u 检验的 II 类风险为 $1-\beta(\mu)=\Phi\left(u_\alpha-\sqrt{n}(\mu-\mu_0)\right)$

当 $\mu>\mu_0$ 时, $1-\beta(\mu)<1-\alpha$, 且 $\mu \to +\infty$ 时, $1-\beta(\mu) \to 0$.

当 I 类风险的显著性水平的 α 减小 (增大)时，II类风险会相应地增大(减小);

并且不论 n 多么大, 只要 n 给定, 总存在 μ_0 附近的点 $\mu(\mu>\mu_0)$ 使 $1-\beta(\mu)$ 几乎等于 $1-\alpha$.

- 只要 $\mu-\mu_0 \geqslant \delta(\delta>0$ 是给定的常数), 则可以确定样本容量 n,使 II 类风险不大于 ε.

例如, 设 $\mu_0=0, \alpha=0.05$, 希望当 $\mu \geqslant 1$ 时,这个 u 检验的 II 类风险不大于 0.10,那么 n 必须满足

$$\Phi\left(u_{0.05}-\sqrt{n}(1-0)\right) \leqslant 0.10,$$
$$u_{0.05}-\sqrt{n} \leqslant -u_{0.10}$$
$$\sqrt{n} \geqslant u_{0.05}+u_{0.10}=1.645+1.282=2.927,$$

即 $n \geqslant 8.57$. 因此,样本容量至少要 9 才能使这个 u 检验的II类风险不大于 0.10

最大功效检验

最优检验评选标准:

在 I 类风险满足显著性水平 α 下, 使 II 风类风险尽可能小, 即要求这个检验的功效函数 $\beta(\theta)$ 满足下述条件:

$$\begin{cases} \beta(\theta) \leqslant \alpha, & \theta \in \Theta_0, \\ \beta(\theta) \text{ 尽可能大}, & \theta \in \Theta_1. \end{cases}$$

给定一个参数型统计问题, 其总体参数 $\theta \in \Theta$, 要检验假设

$$\mathrm{H}_0: \theta \in \Theta_0, \mathrm{H}_1: \theta \in \Theta_1.$$

如果存在一个显著性水平 α 的检验 ϕ^*, 使得对于任意一个显著性水平α的检验 ϕ, 均有

$$\beta_{\phi^*}(\theta) \geqslant \beta_\phi(\theta), \ \forall \theta \in \Theta_1,$$

则 ϕ^*称为这个假设检验问题在显著性水平 α 下的**一致最大功效检验 (uniformly most powerful, UMP)**.
当 H_1 为简单假设时, 称ϕ^*为**最大功效检验**.

Neyman-Pearson定理 设总体 X 的概率密度(或概率分布)函数为 $f(x;\theta),\theta\in\Theta=\{\theta_0,\theta_1\},(X_1,X_2,\cdots,X_n)$ 是取自这个总体的一个样本, 要检验简单假设

$$\mathrm{H}_0:\theta=\theta_0,\ \mathrm{H}_1:\theta=\theta_1.$$

给定显著性水平 α, 如果存在临界值 k,使

$$P_{\theta_0}\left\{\frac{\prod_{i=1}^{n} f(X_i;\theta_1)}{\prod_{i=1}^{n} f(X_i;\theta_0)}\geq k\right\}=\alpha$$

那么,以 $W_i^*=\left\{(x_1,x_2,\cdots,x_n)\mid\frac{\prod_{i=1}^{n} f(x_i;\theta_1)}{\prod_{i=1}^{n} f(x_i;\theta_0)}\geq k\right\}$ 为拒绝域的检验 ϕ^* 是该假设检验问题在显著性水平 α 下的最大功效检验.

似然比 $l(X_1,X_2,\cdots,X_n)=\frac{\prod_{i=1}^{n} f(X_i;\theta_1)}{\prod_{i=1}^{n} f(X_i;\theta_0)}=\frac{L(\theta_1)}{L(\theta_0)}$

似然函数 $L(\theta)$ 可以看成当参数取 θ 值时, 样本 $(X_1,X_2,\cdots,X_n)$ 落在 $(x_1,x_2,\cdots,x_n)$ 的邻域内（或取（$x_1,x_2,\cdots,x_n$））的可能性大小的一个度量。当 H_0 成立即 $\theta=\theta_0$ 时, $l(x_1,x_2,\cdots,x_n)$ 的值应该较小；反之，如果发现 $l(x_1,x_2,\cdots,x_n)$ 较大时,就有理由认为 H_0 可能不成立.

证明：只证明连续总体的情况，离散总体类似。

设 $W = W_i^* = \left\{(x_1, x_2, \cdots, x_n) \mid \frac{\prod_{i=1}^n f(x_i;\theta_1)}{\prod_{i=1}^n f(x_i;\theta_0)} \geq k\right\}$,

$L(x_{1:n};\theta) = \prod_{i=1}^n f(X_i;\theta)$,

$$P_\theta(X_{1:n} \in W) = \int_W L(x_{1:n};\theta)dx_{1:n}$$

对任意满足$P_{\theta_0}(X_{1:n} \in W') \leq \alpha$的拒绝域 W'，有

$$\begin{aligned}
P_{\theta_1}(X_{1:n} \in W) - P_{\theta_1}(X_{1:n} \in W') &= \int_W L(x_{1:n};\theta_1)dx_{1:n} - \int_{W'} L(x_{1:n};\theta_1)dx_{1:n} \\
&= \int_{W-W'} L(x_{1:n};\theta_1)dx_{1:n} - \int_{W'-W} L(x_{1:n};\theta_1)dx_{1:n} \\
&\geq k\left(\int_{W-W'} L(x_{1:n};\theta_0)dx_{1:n} - \int_{W'-W} L(x_{1:n};\theta_0)dx_{1:n}\right) \\
&= k\left(\int_W L(x_{1:n};\theta_0)dx_{1:n} - \int_{W'} L(x_{1:n};\theta_0)dx_{1:n}\right) \\
&= k(\alpha - P_{\theta_0}(X_{1:n} \in W')) \geq 0.
\end{aligned}$$

这表明W的功效不小于W'，根据UMP的定义，得证。

例 设总体分布为离散型, 其概率函数 $f(x;\theta)$ 由下表给出, 其中 $\theta \in \{0,1\}$.

θ \ x ($f(x;\theta)$)	0	1	2
0	0.8	0.1	0.1
1	0.2	0.6	0.2

X_1 是取自这个总体的大小为 1 的样本, 要检验 $\mathrm{H}_0\colon\theta = 0,\ \mathrm{H}_1\colon\theta = 1$.

$$\text{似然比 } l(x_1) = \begin{cases} 0.25, & x_1 = 0, \\ 6, & x_1 = 1, \\ 2, & x_1 = 2. \end{cases}$$

在显著性水平 $\alpha = 0.1$ 下, 因为 $P_{\theta=0}\{l(x_1) \geq 6\} = 0.1$，最大功效检验 ϕ^* 给出的拒绝域为 $W_1^* = \{x_1 \mid x_1 = 1\}$, 这个检验在 $\theta = 1$ 处的功效为 $\beta_{\phi^*}(1) = P_{\theta=1}\{X_1 \in W_1^*\} = P_{\theta=1}\{X_1 = 1\} = 0.6$. 这个检验的I类风险为 0.1, II 类风险为$1 - 0.6 = 0.4$

考虑另一个检验 ϕ, 它的拒绝域为 $W_1 = \{x_1 \mid x_1 = 2\}$. 这个检验的I类风险仍为0.1, II类风险为$1 - \beta_\phi(1) = 1 - P_{\theta=1}\{X_1 = 2\} = 1 - 0.2 = 0.8$

例 设 $(X_1, X_2, \cdots, X_n)$ 是取自正态总体 $N(\mu; 1)$ 的一个样本,其中 μ 未知,要检验

$$\mathrm{H}_0: \mu = \mu_0,\ \mathrm{H}_1: \mu = \mu_1.$$

其中 $\mu_0 < \mu_1$. 在显著性水平 α 下, 求最大功效检验的拒绝域.

解 $\prod_{i=1}^{n} f(x_i; \mu_1) = (2\pi)^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu_1)^2\right]$, $\prod_{i=1}^{n} f(x_i; \mu_0) = (2\pi)^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu_0)^2\right]$,

$l(x_1, x_2, \cdots, x_n) = \exp\left[n(\mu_1 - \mu_0)\bar{x} - \frac{n}{2}\left(\mu_1^2 - \mu_0^2\right)\right]$.

由Neyman-Pearson定理, 最大功效检验的拒绝域为

$W_1^* = \{(x_1, x_2, \cdots, x_n) \mid l(x_1, x_2, \cdots, x_n) > k\} = \{(x_1, x_2, \cdots, x_n) \mid \bar{x} > k_1\}$
$= \left\{(x_1, x_2, \cdots, x_n) \mid \sqrt{n}(\bar{x} - \mu_0) > k_0\right\}$

其中 $k_1 = \left[\ln k + \frac{n}{2}\left(\mu_1^2 - \mu_0^2\right)\right] / n(\mu_1 - \mu_0)$, $k_0 = \sqrt{n}(k_1 - \mu_0)$.

当 H_0 成立, 即 $\mu = \mu_0$ 时, $\sqrt{n}(\bar{X} - \mu_0) \sim N(0; 1)$. 因此,取临界值 $k_0 = u_\alpha$ 便有

$$P_{\mu_0}\left\{\sqrt{n}(\bar{X} - \mu_0) > u_\alpha\right\} = \alpha.$$

最大功效检验的拒绝域 $W_1^* = \left\{(x_1, x_2, \cdots, x_n) \mid \sqrt{n}(\bar{x} - \mu_0) > u_\alpha\right\}$.

例 设 $(X_1, X_2, \cdots, X_n)$ 是取自 $X \sim N(\mu; 1)$ 的一个样本,其中 μ 末知. 证明对于单侧假设检验问题(1) $\mathrm{H}_0: \mu \leqslant \mu_0$, $\mathrm{H}_1: \mu > \mu_0$, 拒绝域为 $W_1^* = \left\{(x_1, x_2, \cdots, x_n) \mid \sqrt{n}(\bar{x} - \mu_0) > u_\alpha\right\}$ 的 u 检验 ϕ^* 是显著性水平 α 下的一致最大功效检验.

证 对任一固定的 $\mu_1 > \mu_0$, 对于检验问题(2)

$$\mathrm{H}_0: \mu = \mu_0,\ \mathrm{H}_1: \mu = \mu_1,$$

由上例知, ϕ^* 是显著性水平 α 下的最大功效检验. 由于 ϕ^* 的拒绝域 $W_1^* = \left\{(x_1, x_2, \cdots, x_n) \mid \sqrt{n}(\bar{x} - \mu_0) > u_\alpha\right\}$ 与 μ_1 的取值无关,所以 ϕ^*也是检验问题(3) $\mathrm{H}_0: \mu = \mu_0$, $\mathrm{H}_1: \mu > \mu_0$ 的一致最大功效检验.

设 ϕ 是检验问题(1) $\mathrm{H}_0: \mu \leqslant \mu_0$, $\mathrm{H}_1: \mu > \mu_0$ 的任意一个显著性水平 α 下的检验,它的拒绝域为 W_1. 由显著性水平 α 的要求知,当 $\mu \leqslant \mu_0$ 时,

$$P_\mu\{(X_1, X_2, \cdots, X_n) \in W_1\} \leqslant \alpha,$$

因而 ϕ 也是假设检验问题(3)的一个显著性水平 α 下的检验. 因为 ϕ^* 是检验问题(3)的一致最大功效检验, 所以对任一 $\mu > \mu_0$, 有 $\beta_{\phi^*}(\mu) \geqslant \beta_\phi(\mu)$.

因此ϕ^*是假设检验问题(1)的一致最大功效检验.

可以类似地证明，前面推导的仅含一个总体参数的单侧检验都是一致最大功效检验.

填空题 3分

设置

假设随机变量 X 的概率密度函数为 $f(x)=(1+\theta)x^{\theta}, 0<x<1$, 现考虑假设检验问题:

$$H_0: \theta=5, H_1: \theta=3$$

拒绝域为 $\{X>1/2\}$.

该检验问题的I类错误概率= [填空1]，II类错误概率= [填空2]，

$\theta=2$ 时功效函数值= [填空3] （答案以分数a/b形式表示）

1. I类错误概率 (α)

I类错误概率为:

$$\alpha=P(X>\frac{1}{2}\mid H_0)$$

在 H_0 下，$\theta=5$，因此:

$$f(x;5)=6x^5,\quad 0<x<1$$

计算:

$$\alpha=\int_{1/2}^{1}6x^5dx=[x^6]_{1/2}^{1}=1^6-\left(\frac{1}{2}\right)^6=1-\frac{1}{64}=\frac{63}{64}$$

功效函数定义

对于假设检验问题:

- 原假设 $H_0:\theta=\theta_0$
- 备择假设 $H_1:\theta\neq\theta_0$

功效函数记为 $\beta(\theta)$，定义为:

$$\beta(\theta)=P(\text{拒绝 } H_0\mid\theta)$$

即在给定参数值 θ 时，检验拒绝 H_0 的概率。

- 犯第一类错误的概率为

$P\{\text{拒绝 } H_0 \mid H_0\text{成立}\}$
$=P\{(X_1,X_2,\cdots,X_n)\in W_1 \mid H_0\text{成立}\}$

称为检验的 I 类风险;

- 犯第二类错误的概率为

$P\{\text{接受}H_0 \mid H_0\text{不成立}\}$
$=P\{(X_1,X_2,\cdots,X_n)\notin W_1 \mid H_0\text{不成立}\}$

称为检验的 II 类风险.

2. II类错误概率 (β)

II类错误概率为:

$$\beta=P(X\leq\frac{1}{2}\mid H_1)$$

在 H_1 下，$\theta=3$，因此:

$$f(x;3)=4x^3,\quad 0<x<1$$

计算:

$$\beta=\int_0^{1/2}4x^3dx=[x^4]_0^{1/2}=\left(\frac{1}{2}\right)^4-0^4=\frac{1}{16}$$

3. 功效函数值在 $\theta=2$ 时

功效函数值为:

$$P(X>\frac{1}{2}\mid\theta=2)$$

在 $\theta=2$ 时:

$$f(x;2)=3x^2,\quad 0<x<1$$

计算:

$$P(X>\frac{1}{2}\mid\theta=2)=\int_{1/2}^{1}3x^2dx=[x^3]_{1/2}^{1}=1^3-\left(\frac{1}{2}\right)^3=1-\frac{1}{8}=\frac{7}{8}$$

作答