第 七 章

假 设 检 验

Hypothesis Testing

假设检验 vs 参数估计

- 若对参数一无所知，则用参数估计方法处理。
- 若对参数有怀疑，需要证实，则用假设检验方法处理。
- 假设检验：

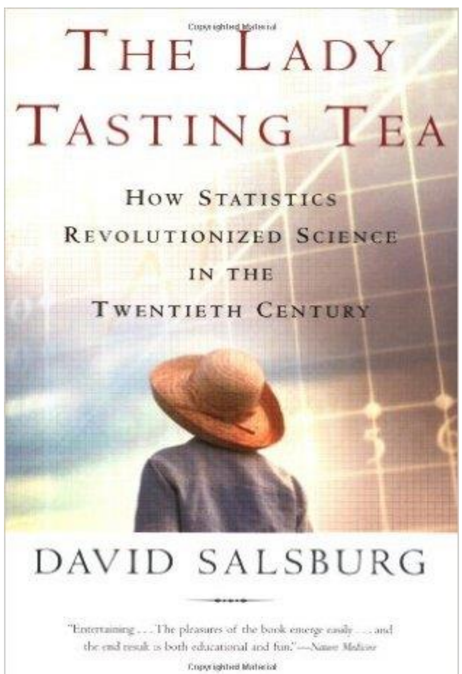
  用来推断样本与总体，样本与样本的差异是由

  抽样误差 引起的，还是 本质差别 造成的。

第一节:基本思想与步骤

例（女士品茶）：

某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），也可以先倒奶后倒茶（记为MT）。某女士声称她可以辨别是TM型，还是MT型。可信吗？



第一节:基本思想与步骤

著名统计学家Fisher建议做实验来检验假设(命题):

Null Hypothesis H_0：**该女士无此种鉴别能力(p=0.5)。**

Alternative Hypothesis H_1：**该女士有此种鉴别能力(p>0.5).**

方法:

准备若干杯（比如10杯）奶茶(TM or MT),让她品尝鉴别。

如果她正确地判别出 A 杯（比如10杯，或8杯),

如何判断：该女士究竟有无此种鉴别能力?

我们是否相信该女士有此种鉴别能力?

第一节:基本思想与步骤

Null Hypothesis H_0：该女士无此种鉴别能力。

Alternative Hypothesis H_1: 该女士有此种鉴别能力。

- How likely are we to observe data similar to the given data set, assuming H_0 is correct?
- If the observed data is unlikely to occur under H_0, then we decide that its logic opposition H_1 is true.

原假设H0：该女士无此种鉴别能力(p=0.5)

没有任何先验知识（比如不知10杯里有几杯TM，几杯MT）.

◆ 如果 A = 10，该女士究竟有无此种鉴别能力？

分析：

- 如果该女士无此种鉴别能力，她只能猜，每次猜对的概率为 0.5. 10次都对的概率 $0.5^{10} < 0.001$，很小。
- 小概率事件在一次实验中几乎不会发生。
- 现在居然发生了，只能说明原假设不当，即应当认为该女士有此种鉴别能力。

We just observed an unlikely event, an event so unlikely that we must doubt the assumption. So we have strong evidence to reject H0.

原假设H0：该女士无此种鉴别能力(p=0.5)

- 如果 A = 8，该女士有此种鉴别能力吗？

  即如何对假设H0作出判断？

用二项分布 B(10,0.5)计算相应的概率。

能猜对 8杯或更多杯 的概率为：0.055.

对原假设不利

1-pbinom(7, 10, 0.5)
= 0.0546875

0.05 < 概率 < 0.1:

weak evidence against H0.

原假设H0：该女士无此种鉴别能力(p=0.5)

有一些信息：设10杯奶茶中有5杯TM，另5杯MT。

◆ 如果 A = 10，该女士究竟有无此种鉴别能力？

分析：

- 如果她无此种鉴别能力,她只能随机分成两组:TM组和MT组。
- 总分法数量：$C_{10}^{5} = 252.$
- 只有一种是正确的。正确的概率为 1/252=0.004.很小。
- 小概率事件在一次实验中几乎不会发生。
- 现在居然发生了，只能说明原假设不当，即应当认为该女士有此种鉴别能力。

We just observed an unlikely event, an event so unlikely that we must doubt the assumption. So we have the reason to reject H0.

原假设H_0：该女士无此种鉴别能力(p=0.5)

有一些信息：设10杯奶茶中有5杯TM，另5杯MT。

- 如果A = 8,
  - 总分法数量： $C_{10}^{5} = 252.$
  - 分对8杯的组合数：TM型正确选了4杯（5选4），MT型正确选了4杯（即从5杯MT中误选了1杯作为TM），共有5*5=25种可能，概率 25/252 = 0.099.
  - 所以，(在无鉴别能力的情况下)，能 猜对8杯或更多杯的概率为 (1+25)/252 = 0.103.

（不利于H0的）

概率 > 0.1:

No evidence against H_0 (H_0 is not rejected).

原假设H0：该女士无此种鉴别能力(p=0.5)

方法一： 在假设H_0为真的条件下，计算出现所观察到的情况或更极端情况（更不利于H_0）的概率（即p值）。

① 如果这个概率很小（比如小于0.05），则拒绝假设H_0，即说明她有这种能力。

② 如果这个概率不小（比如大于0.1），则不能拒绝假设H_0，或没有充分的证据拒绝假设H_0.

方法二： 用某种方法确定"临界杯数"， （临界杯数如何确定？）

① 如果实际观测得到的正确杯数大于或等于"临界杯数"，则拒绝假设H_0，即说明她有这种能力。

② 如果实际观测得到的正确杯数小于"临界杯数"，则没有充分的证据拒绝假设H_0.

问题：

- 如何设计实验？
- 如何计算出现所观察到的情况或更极端（更不利于H_0）情况的概率？ （p值）
- 如何确定“临界杯数”？ （拒绝域）

- 判断会发生错误吗？
- 发生错误的概率是多少？
- 如何控制犯错误的概率？

一、假设检验问题

例： 某厂生产的合金强度服从 $N(\theta,16)$，其中θ 的设计值不低于110(Pa). 某天随机抽取 25 个样本，测得强度为 x_1，…，x_{25}，均值为108（Pa）. **问当日生产是否正常？**

命题：“合金平均强度不低于110(Pa)”--- 假设检验问题。

两个统计假设：

$$H_0: \theta \in \Theta_0 = \{\theta : \theta \geq 110\} \quad \text{vs} \quad H_1: \theta \in \Theta_1 = \{\theta : \theta < 110\}$$

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

判断假设(命题)H_0是否成立：“检验”. **如何判断？**

检验结果：“原假设不正确”--- 拒绝 H_0；

没理由认为“原假设不正确”--- 不能拒绝 H_0.

二、基本步骤

1. 建立假设

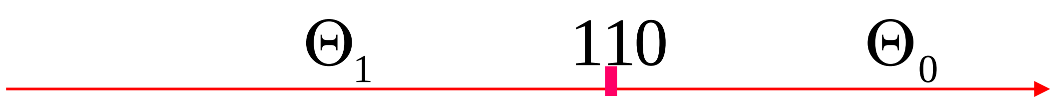
原假设H_0： 被检验的假设，通常**是受保护而不轻易否定**的假设。

备择假设H_1： 希望证明的假设 (the alternative is the claim that we wish to prove).

对上例：

$$H_0: \theta \in \Theta_0 = \{\theta : \theta \geq 110\} \quad \textbf{vs} \quad H_1: \theta \in \Theta_1 = \{\theta : \theta < 110\}$$

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



2. 选择检验统计量，给出拒绝域的形式

检验统计量： 由样本构造的统计量，用于对原假设进行判断。

拒绝域： 使原假设被拒绝的样本观测值所在区域。

考虑样本均值 $\overline{x}$: 样本均值是总体均值θ的一个估计；

样本均值提供总体均值的信息。

样本均值$\overline{x}$越大$\xleftrightarrow{\text{常常}}$总体均值$\theta$也大（$H_0$为真）；

样本均值$\overline{x}$越小$\xleftrightarrow{\text{常常}}$总体均值$\theta$也小（$H_1$为真）；

如果$\overline{x}$过分小(小于某临界值c),则有理由怀疑H_0的真实性。

所以，合理的拒绝域应为：

$W := \{(x_1,\ldots,x_n):\overline{x} \le c\} = \{\overline{x} \le c\}$，其中$c$为待定常数(临界值).

结论：如果$(x_1,\ldots,x_n) \in W$,则认为H_0不成立(即拒绝H_0);

如果$(x_1,\ldots,x_n) \notin W$,则认为$H_0$成立(即接受$H_0$);

3. 选择“显著性水平”－ 控制犯“弃真”错误的概率

检验结果可能与实际吻合，或不吻合，即可能犯错误. 何种错误?

真实情况 / 所作判断	H_0为真	H_1为真
接受H_0 $(x_1,\ldots,x_n)\notin W$	正确 ☺ ✓	取伪 (犯第II类错误)
拒绝H_0 $(x_1,\ldots,x_n)\in W$	弃真 (犯第I类错误)	正确 ☺ ✓

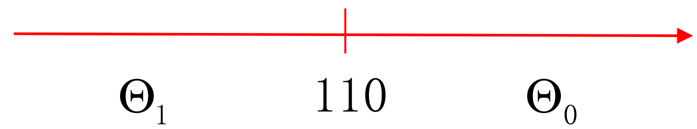
思想：控制犯第I类错误的概率。

Level of significance is the prob of making Type I error

$$P(\text{拒绝}H_0\,|\,\text{当}H_0\text{为真})\le\alpha,$$

α：显著性水平(事先给定，一般选为0.05,0.01等)

4. 决定拒绝区域



希望：犯第I类错误的概率 $P(\text{拒绝}H_0 \mid \text{当}H_0\text{为真}) \le \alpha$，

希望由此决定临界值

当 $\theta \in \Theta_0 = \{\theta \ge 110\}$ 时，

这个概率有多大？最大有多大？

$$P(\text{拒绝}H_0 \mid \text{当}H_0\text{为真}) = P_\theta\{(X_1,\dots,X_n)\in W\} = P_\theta(\overline{X} \le c)$$

$$= P_\theta\left(\frac{\overline{X}-\theta}{\sigma/\sqrt{n}} \le \frac{c-\theta}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{c-\theta}{\sigma/\sqrt{n}}\right) \le \Phi\left(\frac{c-110}{\sigma/\sqrt{n}}\right).$$

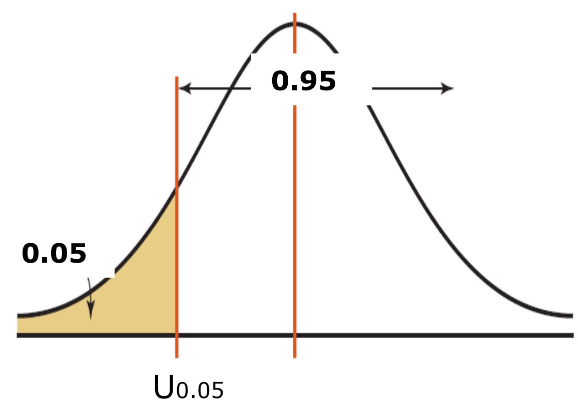
关于Theta 单调减
H0 为真时，Theta
的最小可能为110

若要 $P_\theta\{(X_1,\dots,X_n)\in W\} \le \alpha$，只要 $\Phi\left(\frac{c-110}{\sigma/\sqrt{n}}\right) = \alpha$. 由此可确定 c.

对 $\alpha = 0.05$，$\sigma = 4$，$n = 25$，

有 $\frac{c-110}{4/5} = u_{0.05} = -u_{0.95} = -1.64 \Rightarrow c = 108.684.$

拒绝域为：$W = \{\overline{X} \le 108.684\}.$



5. 作出判断

拒绝域为：$W=\{\overline{X}\le 108.684\}$.

* 当$\overline{x}\le 108.684$时，拒绝H_0;

  因: $P(\overline{X}\le 108.684$, 当$H_0$为真$)\le\alpha$,是小概率事件,

  如果居然发生了,则拒绝H_0.

* 当$\overline{x}>108.684$时，接受H_0

  (未导致小概率事件发生，没有理由拒绝H_0).

We observed an unlikely event, an event so unlikely that we must doubt the assumption. So we have the reason to reject H0.

例中，$\overline{x}=108\le 108.684$，拒绝$H_0$，即认为该日生产不正常。

如果对相同样本容量n，测得$\overline{x}=109$，则不能拒绝原假设，即接受H_0.

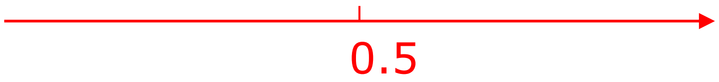
例：

糖果包装机，包出的糖果重量为RV，服从正态分布。
正常时，均值为 0.5 kg，标准差 0.015 kg. 某日开工，
抽取 9 袋，算得样本均值为 0.511 kg. 问机器是否正常？

建立假设： $H_0: \mu = \mu_0 = 0.5 \quad \text{vs} \quad H_1: \mu \neq \mu_0;$

分析：样本均值$\bar{x}$是μ的估计量。

样本均值提供总体均值的信息。



若H_0为真，$|\bar{x} - \mu_0|$一般不应太大；

若$|\bar{x} - \mu_0|$过分大(大于某临界值)，有理由怀疑H_0的真实性。

可用$|\bar{x} - \mu_0|$的大小或标准化$\dfrac{|\bar{X} - \mu_0|}{\sigma / \sqrt{n}} =: U$后的大小来判断

即：若$|U| \geq c$ (c为某临界值,待定)，则拒绝H_0.

例（续）

(1) 检验统计量：$U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$;

(2) 拒绝域形式：$W=\{(X_1,\dots,X_n):|U|\geq c\}$, c待定.

(3) 给定显著性水平：α $(=0.05$或$0.01,\dots)$. 希望$P($拒绝$H_0|$当H_0为真$)\leq\alpha$.

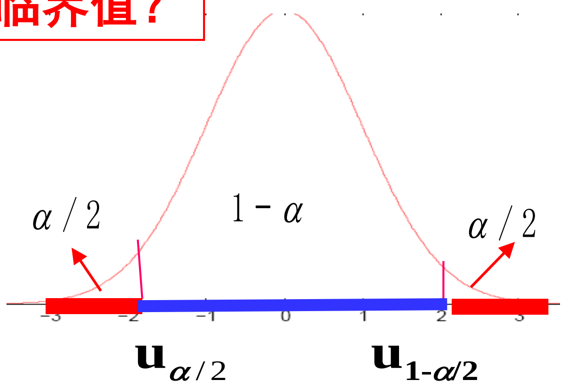
(4) 确定拒绝域：$P($拒绝$H_0|$当H_0为真$)=?$

如何确定临界值？

当H_0为真时,即$\mu=\mu_0$,统计量$U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$.

因 $P($拒绝$H_0|$ 当H_0为真$)=P_{\mu_0}(|U|\geq c)$,

取$c=u_{1-\alpha/2}$,即可使$P_{\mu_0}(|U|\geq c)=\alpha$.



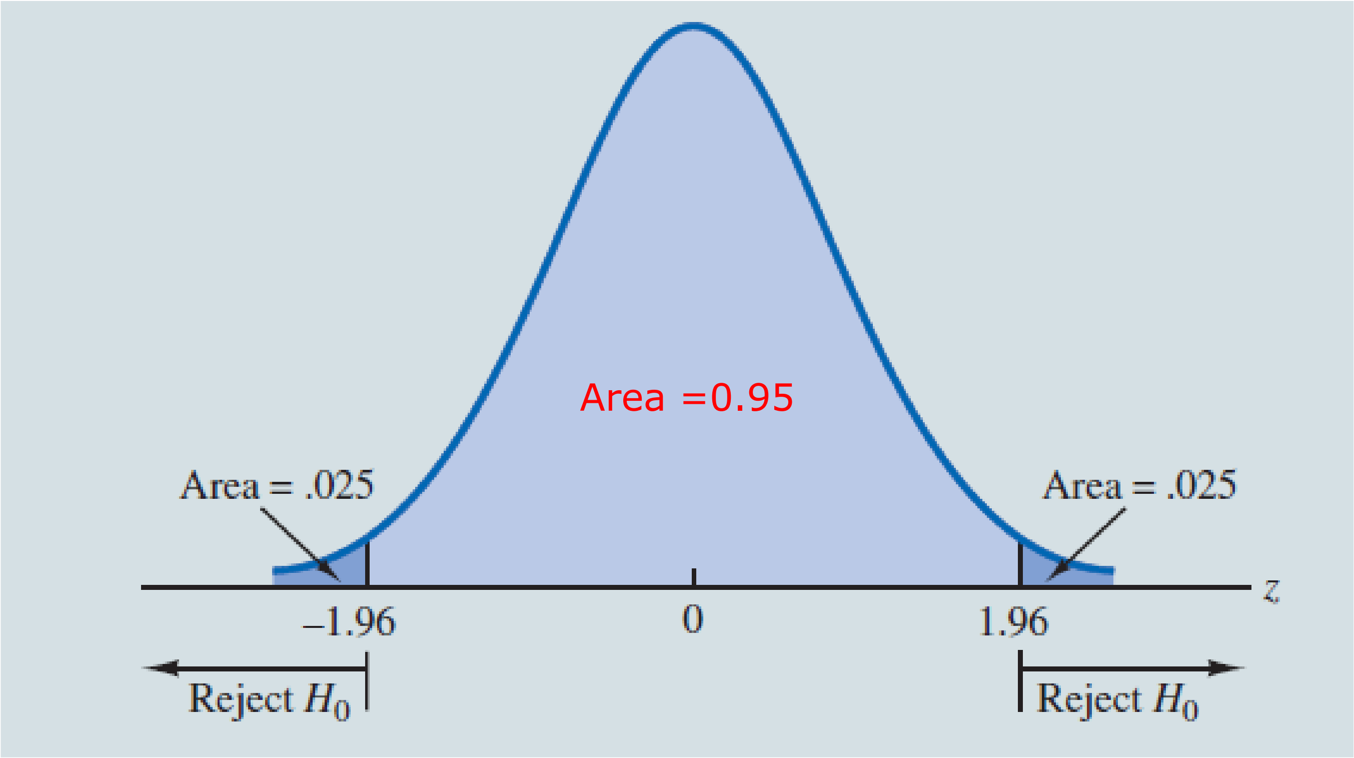
最后的拒绝域为：$W=\{|U|\geq u_{1-\alpha/2}\}$.

当H0为真时，它为小概率事件

(5)现在，取$\alpha=0.05$. $u_{1-\alpha/2}=u_{0.975}=1.96$. 拒绝域为：$W=\{|U|\geq 1.96\}$.

$$|u|=\frac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}}=2.2>1.96,$$ 落在拒绝域，于是拒绝H_0,即认为机器不正常。

显著性水平0.05



$$\boxed{H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \neq \mu_0}$$

Hypothesis Testing - Rejection Region (RR) Method

- **Step 1.** State hypotheses and identify the claim.
- **Step 2.** Find test statistics
- **Step 3.** Find critical value(s) and rejection region.
- **Step 4.** Compute the test value.
- **Step 5.** Make the decision to reject or not reject the null hypothesis.
- **Step 6.** Summarize the results.

注记：

1. “接受H_0”，并不是证明了命题H_0，而只是没有理由拒绝H_0.

In statistics nothing is absolute due to randomness.

2. 比较：

反证法：设“命题”不正确,推得矛盾,所以原“命题”正确。
检验：设H_0为真，如导致“小概率事件”发生，则拒绝H_0.

（不利于H0的）

3. 数学上，我们不能用一个例子去证明一个结论，但可以用一个例子推翻一个结论。

类似，不能用一个样本证明一个假设是成立的，但可以用一个样本推翻一个假设。

Logic of Hypothesis Testing

- We assume that H_0 is true until the sample data demonstrate otherwise.
- This logic is similar to logic of judicial system: presumed innocent until proven guilty.
- We will not be able to determine that a hypothesis statement is true or false.
- We can only assess whether the observed data are consistent with assumption that H_0 is true (*reject* or *fail to reject* a hypothesis).
- If the data would be unlikely when H_0 is true, we reject H_0.

三、犯错误的概率与功效函数（势函数）

所作判断 \ 真实情况	H_0为真	H_1为真
Fail to reject H0 接受H_0 $(x_1,\ldots,x_n)\notin W$	正确 ☺ ✓	取伪（犯第II类错误）
拒绝H_0 $(x_1,\ldots,x_n)\in W$	弃真（犯第I类错误）	正确 ☺ ✓

检验可能犯以下两类错误：

第I类错误：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．（弃真）

第II类错误：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）

三、犯错误的概率与功效函数（势函数）

犯第I类错误的概率$\alpha = P(\text{拒绝}H_0, \text{当}H_0\text{为真})$

（弃真）
$$= P_{\theta\in\Theta_0}(\underline{\text{拒绝}H_0})$$

犯第II类错误的概率$\beta = P(\text{接受}H_0, \text{当}H_1\text{为真})$

（取伪）
$$= 1 - P(\text{拒绝}H_0, \text{当}H_1\text{为真})$$
$$= 1 - P_{\theta\in\Theta_1}(\underline{\text{拒绝}H_0})$$

$\forall \theta \in \Theta_0 \cup \Theta_1$，定义

$$g(\theta) := P_\theta(\text{拒绝}H_0) = P_\theta((X_1,\ldots,X_n) \in W)$$

拒绝域

犯两类错误的概率可用同一个函数表示，即功效函数

三、犯错误的概率与功效函数（势函数）

考虑检验问题 $H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_1$，其拒绝域为 W，定义"功效函数"为：样本观测值落在拒绝域内的概率,即

$$g(\theta) := P_\theta\{(X_1, \ldots, X_n) \in W\}, \quad \theta \in \Theta_0 \cup \Theta_1.$$

Theta 的函数

易知有以下关系(对 θ 分段表达):

$$g(\theta) = \begin{cases} P(\text{拒绝}H_0 \mid \text{当}H_0\text{为真}) = \text{犯第I类错误的概率}\alpha(\theta), & \theta \in \Theta_0; \\ P(\text{拒绝}H_0 \mid \text{当}H_1\text{为真}) = 1 - P(\text{接受}H_0 \mid \text{当}H_1\text{为真}) & \\ \qquad = 1 - \text{犯第II类错误的概率}\beta(\theta), & \theta \in \Theta_1; \end{cases}$$

$$\Rightarrow \begin{cases} \text{犯第I类错误的概率}\ \alpha(\theta) = g(\theta), & \theta \in \Theta_0. \\ \text{犯第II类错误的概率}\beta(\theta) = 1 - g(\theta), & \theta \in \Theta_1. \end{cases}$$

犯两类错误的概率
可由势函数给出

检验的功效（Power of Test）

真实情况 所作判断	H_0为真	H_1为真
接受H_0 $(x_1,\ldots,x_n)\notin W$	正确 ☺ ✓ 概率：$1-\alpha$	取伪 （犯第II类错误） 概率：β
拒绝H_0 $(x_1,\ldots,x_n)\in W$	弃真 （犯第I类错误） α = 显著性水平	正确 ☺ ✓ power of test $1-\beta$

- **Power of Test：$1-\beta$.**

It is prob not making Type II error,
i.e., prob of rejecting a false H_0 --- （正确地拒绝）
or ability of recognizing correctly that H_0 is false.

Remark

- 显著性水平（Significance level，犯第I类错误的概率)：

  α = P(拒绝H_0|H_0)　　（弃真）

- 犯第II类错误的概率(Prob of Type II Error):

  β = P(接受H_0|H_1)　　（取伪）

- **检验的功效**（Power of Test）：

  1−β = P（拒绝H_0|H_1)　　（正确判断H0为F）

- Type I error --- the error of rejecting a true H_0
- Type II error --- the error of failing to reject a false H_0

例（续）

对第一例中的检验问题$H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_1$，设其拒绝域为W，

则势函数：$g(\theta) = P_\theta\{(X_1, \ldots, X_n) \in W\} = P_\theta\{\overline{X} \le c\}, \quad \theta \in \Theta_0 \cup \Theta_1.$

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

这个势函数是
Theta的减函数

$$\Rightarrow \begin{cases} \text{犯第I类错误的概率 } \alpha(\theta) = g(\theta) = \Phi\left(\dfrac{c-\theta}{4/5}\right), & \theta \in \Theta_0. \\ \text{犯第II类错误的概率 } \beta(\theta) = 1 - g(\theta) = 1 - \Phi\left(\dfrac{c-\theta}{4/5}\right), & \theta \in \Theta_1. \end{cases}$$

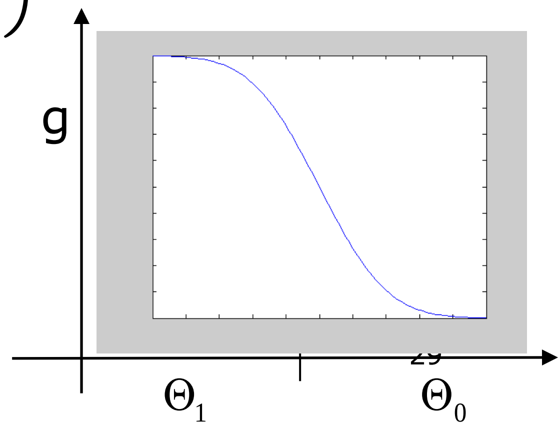
（弃真）

（取伪）

$\alpha(\theta)$小 $\Leftrightarrow c$小 $\Leftrightarrow \beta(\theta)$大。

(减小一类错误的概率会增大另一类错误的概率)

样本量一定时，不可能找到一个使α 和β 都小的检验。



例： 设$x_1,\ldots,x_{10}$设来自$0-1$总体$B(1,p)$的样本，考虑检验问题：

$$H_0: p=0.2 \qquad vs \qquad H_1: p=0.4,$$

取拒绝域为$W=\{\overline{X}\geq 0.5\}$.求该检验犯第$\mathbf{I},\mathbf{II}$类错误的概率。

解： 势函数为 $g(p)=\mathbf{P}_p(\overline{X}\geq 0.5),\ p=0.2$或$0.4$.

犯第I类错误的概率

$$\alpha=g(0.2)=\mathbf{P}_{0.2}(\overline{X}\geq 0.5)=\mathbf{P}_{0.2}(10\overline{X}\geq 5)$$

$$=\sum_{k=5}^{10}C_{10}^k 0.2^k\cdot 0.8^{10-k}=0.03.$$ 注：$10\overline{X}=X_1+\cdots+X_{10}\sim B(10,p)$.

犯第II类错误的概率

$$\beta=1-g(0.4)=1-\mathbf{P}_{0.4}(\overline{X}\geq 0.5)=\mathbf{P}_{0.4}(\overline{X}<0.5)$$

$$=\mathbf{P}_{0.4}(10\overline{X}<5)=\sum_{k=0}^{4}C_{10}^k 0.4^k\cdot 0.6^{10-k}=0.63.$$

拒绝域 --> 势函数 --> 犯两类错误的概率

第二节：正态总体参数假设检验

设$x_1,\ldots,x_n$是来自正态总体$N(\mu,\sigma^2)$.

*关于均值有以下三种假设检验：

(1) $H_0:\mu\geq\mu_0$ **vs** $H_1:\mu<\mu_0$;

(2) $H_0:\mu\leq\mu_0$ **vs** $H_1:\mu>\mu_0$;

(3) $H_0:\mu=\mu_0$ **vs** $H_1:\mu\neq\mu_0$;

*关于方差有以下三种假设检验：

(1) $H_0:\sigma^2\geq\sigma_0^2$ **vs** $H_1:\sigma^2<\sigma_0^2$;

(2) $H_0:\sigma^2\leq\sigma_0^2$ **vs** $H_1:\sigma^2>\sigma_0^2$;;

(3) $H_0:\sigma^2=\sigma_0^2$ **vs** $H_1:\sigma^2\neq\sigma_0^2$;

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-Tailed Test）.

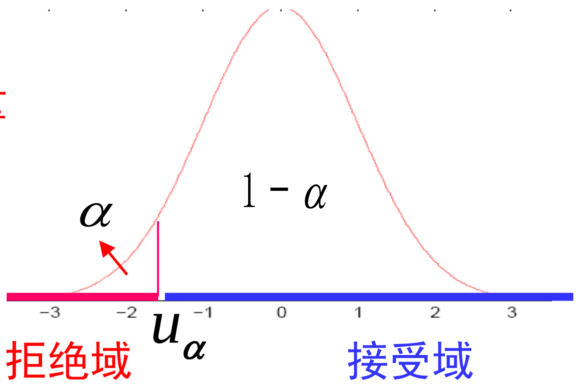
当备择假设 H_1 分散在原假设H_0两侧时的检验称为双侧检验（Two-tailed test）.

一、关于均值的检验（方差已知）

设$x_1,\ldots,x_n$是来自正态总体$N(\mu,\sigma^2)$.考虑检验：

1. $H_0:\mu\ge\mu_0$ vs $H_1:\mu<\mu_0$;

(1) 检验统计量：$U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$; 标准化的检验统计量



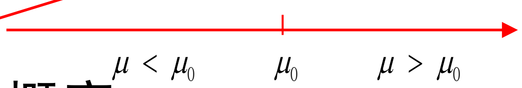
(2) 拒绝域形式：$W=\{(X_1,\ldots,X_n):U\le c\}$, c待定.

(3) 给定显著性水平：α $(=0.05$或$0.01,\ldots)$.

势函数$g(\mu)=P_\mu(U\le c)$

$$=P_\mu\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\le c\right)=P_\mu\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\le\frac{\mu_0-\mu}{\sigma/\sqrt{n}}+c\right)=\Phi\left(\frac{\mu_0-\mu}{\sigma/\sqrt{n}}+c\right)$$是μ的减函数；

(4) 确定拒绝域：



若要 $g(\mu)=P_\mu(U\le c)\le\alpha,\forall\mu\ge\mu_0$,即犯第I类错误的概率$\le\alpha$,

只要$g(\mu_0)=P_{\mu_0}(U\le c)=\alpha$, 即取$c=u_\alpha$ （注：$\mu=\mu_0$时$U\sim N(0,1)$）

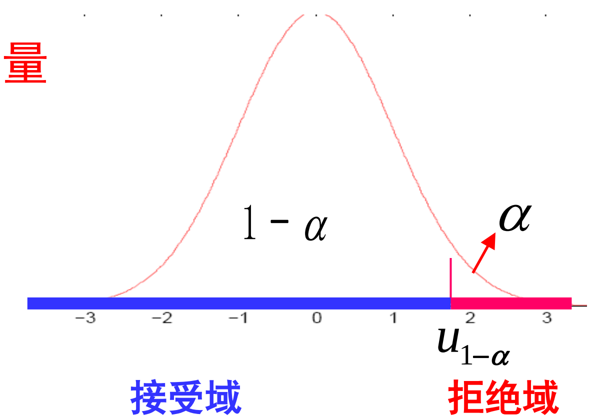
最后的拒绝域为：$W=\{U\le u_\alpha\}$.

犯第一类错误的概率在$\mu=\mu_0$达最大

设$x_1,\dots,x_n$是来自正态总体$N(\mu,\sigma^2)$.考虑检验：

2. $H_0:\mu\le\mu_0 \quad \textbf{vs} \quad H_1:\mu>\mu_0$;

(1) 检验统计量：$U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$;　　标准化的检验统计量



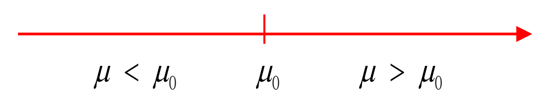
(2) 拒绝域形式：$W=\{(X_1,\dots,X_n):U\ge c\}$, c待定.

(3) 给定显著性水平：α (=0.05或0.01,...).

势函数$g(\mu)=P_\mu(U\ge c)$

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

是μ的增函数；



(4) 确定拒绝域：

若要$g(\mu)=P_\mu(U\ge c)\le\alpha,\forall\mu\le\mu_0$,即犯第I类错误的概率$\le\alpha$,

只要$g(\mu_0)=P_{\mu_0}(U\ge c)=\alpha$，即取$c=u_{1-\alpha}$　($\mu=\mu_0$时,$U\sim N(0,1)$)

最后的拒绝域为：$W=\{U\ge u_{1-\alpha}\}$.

犯第一类错误的概率在$\mu=\mu_0$达最大

设$x_1,\ldots,x_n$是来自正态总体$N(\mu,\sigma^2)$.考虑检验：

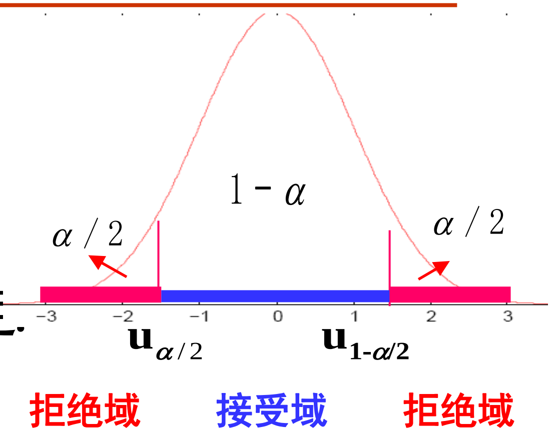
3. $H_0:\mu=\mu_0$ **vs** $H_1:\mu\neq\mu_0$;

(1) 检验统计量 :$U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$;　标准化的检验统计量

(2) 拒绝域形式：$W=\{(X_1,\ldots,X_n):|U|\geq c\}$, c待定.

(3) 给定显著性水平：α $(=0.05或0.01,\ldots)$.

势函数 $g(\mu)=P_\mu(|U|\geq c)$.



(4) 确定拒绝域 :当H_0为真时,即$\mu=\mu_0$,统计量$U\sim N(0,1)$.

因 “犯第I类错误的概率”$=P_{\mu_0}(|U|\geq c)$,

取$c=u_{1-\alpha/2}$就可保证上述概率$=\alpha$

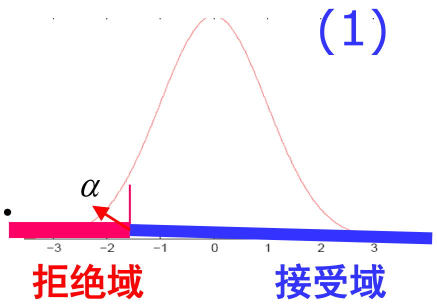
最后的拒绝域为：$W=\{|U|\geq u_{1-\alpha/2}\}$.

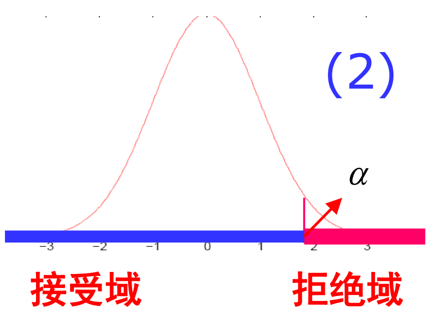
U 检验法总结：给定显著性水平，可以确定拒绝域

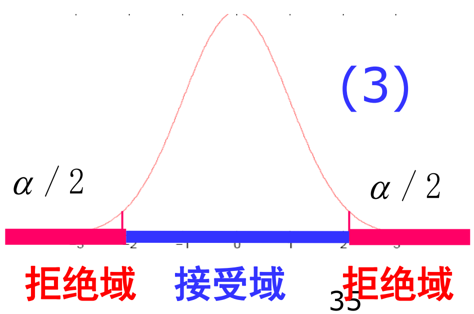
$(1) H_0: \mu \geq \mu_0 \quad$ **vs** $\quad H_1: \mu < \mu_0$；拒绝域：$U \leq u_\alpha$.

$(2) H_0: \mu \leq \mu_0 \quad$ **vs** $\quad H_1: \mu > \mu_0$；拒绝域：$U \geq u_{1-\alpha}$.

$(3) H_0: \mu = \mu_0 \quad$ **vs** $\quad H_1: \mu \neq \mu_0$；拒绝域：$|U| \geq u_{1-\alpha/2}$.







对三种假设检验问题：

* 检验统计量均为：$U = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$；

* 都只要在$\mu = \mu_0$控制犯第I类错误的概率$= \alpha$即可；

* 在$\mu = \mu_0$时，$U = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$.

由P_{μ_0}(样本落在拒绝域)$= \alpha \Rightarrow$ 可确定临界值。

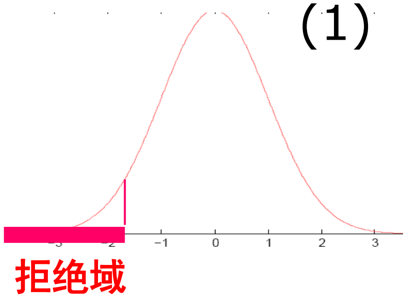
(拒绝域的形式要看备择假设)

U 检验法总结：

$(1) H_0: \mu \geq \mu_0 \quad \text{vs} \quad H_1: \mu < \mu_0$； 拒绝域：$U \leq u_\alpha$.

$(2) H_0: \mu \leq \mu_0 \quad \text{vs} \quad H_1: \mu > \mu_0$； 拒绝域：$U \geq u_{1-\alpha}$.

$(3) H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \neq \mu_0$； 拒绝域：$|U| \geq u_{1-\alpha/2}$.

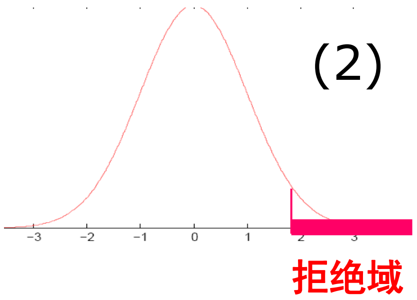


(1)

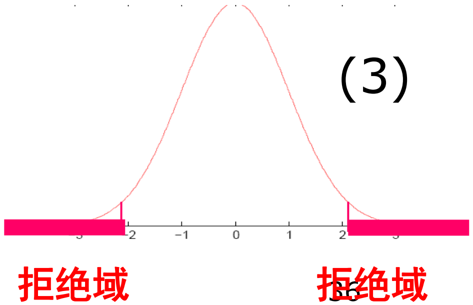
前两者可变形：

$(1)^* \ H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu < \mu_0$； 拒绝域：$U \leq u_\alpha$.

$(2)^* \ H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu > \mu_0$； 拒绝域：$U \geq u_{1-\alpha}$.



(2)



(3)

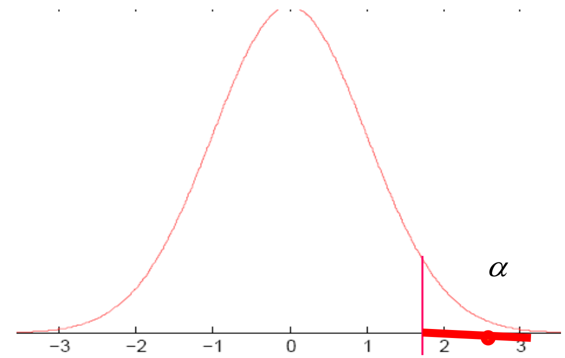
例: 天然牛奶的冰点温度近似服从正态分布,均值-0.545度,标准差0.008度。牛奶掺水可使冰点温度升高而接近水的冰点0度. 现检测 5 批牛奶的冰点温度，均值为-0.535度。

问是否可以认为牛奶中掺了水? 显著性水平 0.05 (或0.01, 0.001).

解：需检验假设

$$H_0: \mu \le \mu_0 = -0.545 \text{ (即设牛奶未掺水)};$$

$$H_1: \mu > \mu_0 \text{ (即设牛奶已掺水)}.$$



$(u_{1-0.01} = 2.33,\ u_{1-0.001} = 3.01)$

其拒绝域为: $U = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \ge u_{1-0.05} = 1.645.$

现在, $u = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{-0.535 - (-0.545)}{0.008 / \sqrt{5}} = 2.80 > 1.645,$

落在拒绝域中，所以在显著性水平$\alpha = 0.05$下拒绝H_0而接受H_1，即认为牛奶已掺水

(在显著性水平α=0.01下，u=2.80>2.33，落在对应的拒绝域中拒绝H_0而接受H_1,即认为牛奶已掺水)

(在显著性水平α=0.001下，u=2.80<3.01，未落在对应的拒绝域中，不能拒绝原假设，故接受H_0)

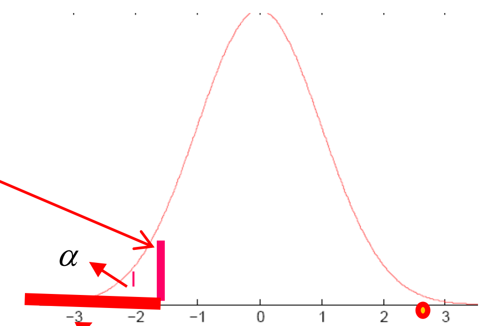
例(续): 天然牛奶的冰点温度近似服从正态分布，均值-0.545度，标准差0.008度。现检测 5 批牛奶的冰点温度，均值为-0.535度。问是否可以认为牛奶中掺了水？显著性水平 0.05.

解：如果牛奶掺水是普遍行为(整个行业失去诚信)，需检验假设

$H_0: \mu \geq \mu_0 = -0.545$ (即设牛奶已掺水);

$H_1: \mu < \mu_0$ (即设牛奶未掺水).

其拒绝域为：$U = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \leq u_{0.05} = -u_{0.95} = -1.645.$



现在，$u = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{-0.535 - (-0.545)}{0.008 / \sqrt{5}} = 2.80 > -1.645,$

不落在拒绝域中,所以在显著性水平0.05下接受H_0,即认为牛奶已掺水

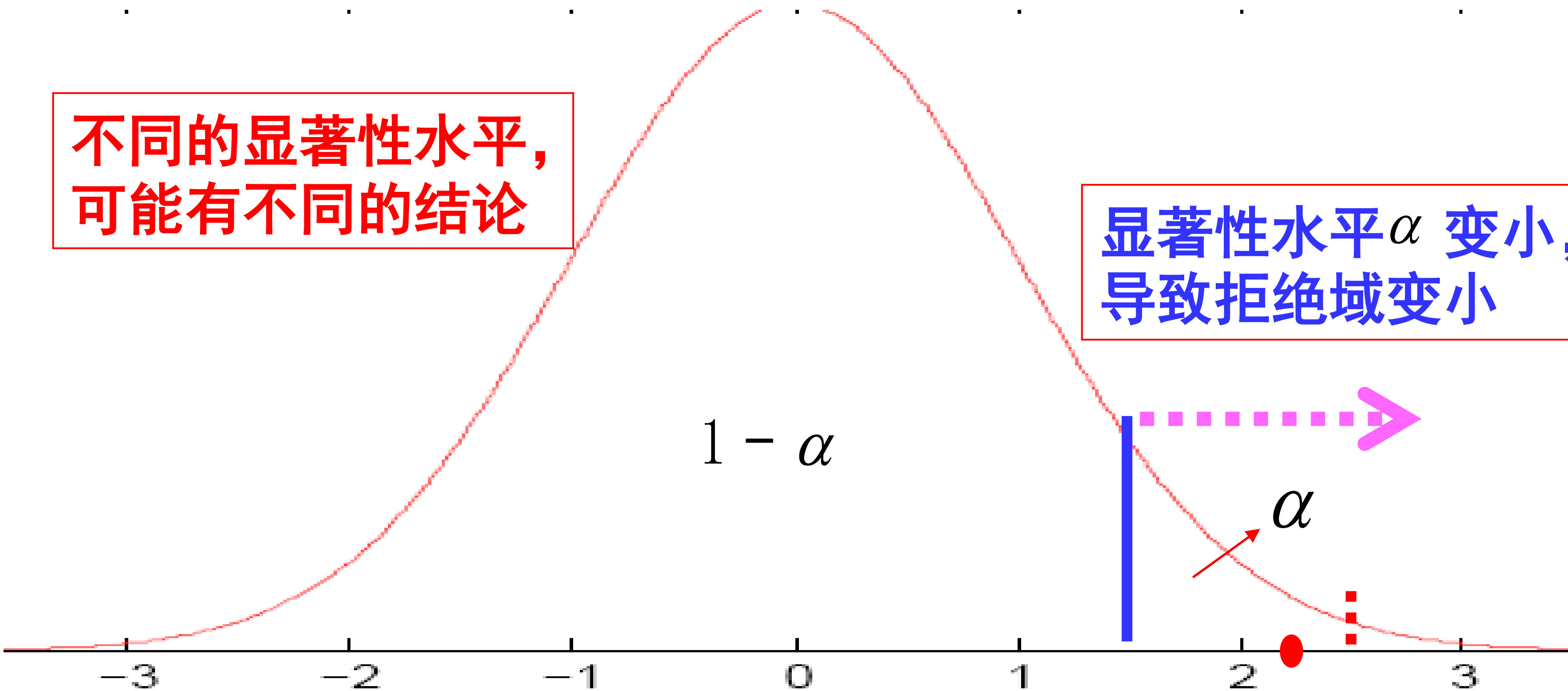
注：当$\overline{x} = -0.545$时,$u = 0 > -1.645$,仍然有相同结论,即认为牛奶已掺水。

当$\overline{x} = -0.555$时,$u = -2.79510 \leq -1.645$,落在拒绝域,即拒绝H_0,认为未掺水

Remark：显著性水平与拒绝域

不同的显著性水平，
可能有不同的结论

显著性水平α 变小，
导致拒绝域变小



U0
（检验统计量观测值）

显著性水平 α 变小，导致拒绝域变小，
原来落到拒绝域中的观测值可能落入接受域。

比较

- 第一种情况:

  $H_0: \mu \le \mu_0 = -0.545$(即设牛奶未掺水)

  $H_1: \mu > \mu_0$(即设牛奶已掺水).

  弃真（第一类错误）：未掺水，判断为已掺水

  取伪（第二类错误）：已掺水，判断为未掺水

- 第二种情况:

  $H_0: \mu \ge \mu_0 = -0.545$(即设牛奶已掺水)

  $H_1: \mu < \mu_0$(即设牛奶未掺水).

  弃真（第一类错误）：已掺水，判断为未掺水

  取伪（第二类错误）：未掺水，判断为已掺水

注：原假设和备择假设的选取

1. 两类假设并非对称。一般选取“维持现状”为原假设。检验设计偏向“原假设”。原假设通常是受保护，而不轻易否定的假设。除非有相当充分的证据，否则维持原假设。

2. 备择假设通常表示所要验证的猜测 (more important). (whatever you are trying to show stat must be represented by the alternative hypothesis)

3. 选择使得两类错误中后果严重的错误成为第Ⅰ类错误(弃真).

犯第Ⅰ类错误的概率受到控制

Example

- A new teaching method is developed that is believed to be better than the current one.

✓ Null hypothesis: new method is no better.

✓ Alternative hypothesis: new method is better.

注记

在司法中（疑罪从无）

H0：被告无罪　　H1：被告有罪

对应的第一类错误（弃真）是什么？

弃真：无罪判有罪（本无罪，却判为X刑）
取伪：有罪判无罪

如果在司法中采用

H0：被告有罪　　H1：被告无罪

对应的第一类错误（弃真）是什么？

弃真：有罪判无罪（有罪，却逍遥法外）
取伪：无罪判有罪（这个概率可能很大）

例:

甲地发讯号μ，乙地接收讯号的值服从$N(\mu,0.2^2)$。现甲地发送同一讯号5次，乙地接收到的讯号值为：$8.05, 8.15, 8.2, 8.1, 8.25$.

问能否认为甲地发送的讯号值为8.(显著性水平为0.05,或0.1).

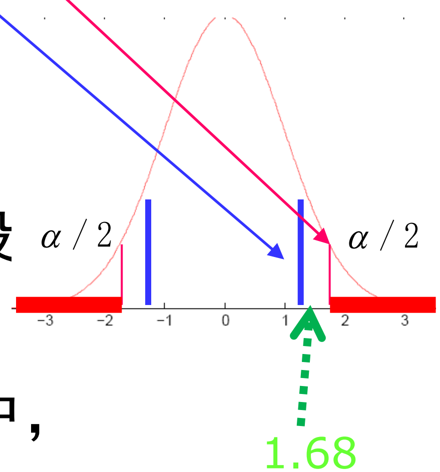
解：需检验假设 $H_0:\mu=8 \quad vs \quad H_1:\mu\neq 8.$

其拒绝域为：$|U|=\left|\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right|\geq u_{1-\alpha/2}=\begin{cases}u_{0.975}=1.96, & \alpha=0.05.\\ u_{0.95}=1.645, & \alpha=0.1.\end{cases}$

现在，$|u|=\left|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|=\left|\dfrac{8.15-8}{0.2/\sqrt{5}}\right|=1.68.$

$*$对显著性水平$\alpha=0.05$, $|u|=1.68<1.96$,

未落在拒绝域中,所以不能拒绝原假设，即接受原假设



$*$对显著性水平$\alpha=0.1$, $|u|=1.68>1.645$, 落在拒绝域中，

拒绝原假设，故接受H_1.

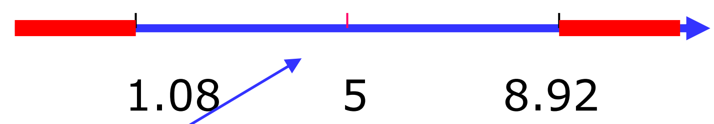
例: 设总体$X \sim N(\mu,4^2)$, $X_1,\ldots,X_4$为来自总体X的样本。

考虑检验问题: $H_0:\mu=5 \quad vs \quad H_1:\mu\neq 5$. 显著性水平为$\alpha=0.05$.

---

(1)求该检验问题的拒绝域;

(2)若总体的均值为$\mu=6$，试计算犯第II类错误("取伪")的概率。



解:(1)检验统计量$U=\dfrac{\overline{X}-5}{4/\sqrt{4}}=\dfrac{\overline{X}-5}{2}$;

拒绝域: $|U|=\left|\dfrac{\overline{X}-5}{2}\right|\geq u_{1-\alpha/2}=u_{0.975}=1.96$.即当$\overline{X}\leq 1.08$或$\overline{X}\geq 8.92$时，拒绝$H_0$.

(2)上述检验问题的接受域为$\overline{X}\in(1.08,8.92)$，要求$\beta:=P_{\mu=6}\{\overline{X}\in(1.08,8.92)\}$.

真值为6，却接受为5（H0）

当$\mu=6$时, $\overline{X}\sim N(6,\dfrac{4^2}{n})$,即$\overline{X}\sim N(6,2^2)$.

$$P_{\mu=6}\{\overline{X}\in(1.08,8.92)\}=P_{\mu=6}\{1.08<\overline{X}<8.92)\}$$

$$=P_{\mu=6}\{\frac{1.08-6}{2}<\frac{\overline{X}-6}{2}<\frac{8.92-6}{2})\}=\Phi(1.46)-\Phi(-2.46)=0.921.$$

Power = 1 – prob of Type II error (beta)

设总体$X \sim N(\mu,2^2), X_1,\ldots,X_{16}$为来自总体$X$的样本。

考虑检验问题：$H_0:\mu=6 \quad vs \quad H_1:\mu\neq 6$.拒绝域取为$W=\{|\overline{X}-6|\geq c\}$.

(1)求c使得该检验问题的显著性水平为0.05;

(2)若总体的均值为$\mu=6.5$，试计算犯第II类错误的概率。

解:(1)当$\mu=6$(即H_0为真)时，$\overline{X}\sim N(6,\ 4/16)$,即$\overline{X}\sim N(6,\ 0.5^2)$

显著性水平 $0.05=P_{\mu=6}(|\overline{X}-6|\geq c)$ (犯第I类错误的概率)

$$\Rightarrow 0.025=P_{\mu=6}(\overline{X}-6\geq c)=P_{\mu=6}(\frac{\overline{X}-6}{0.5}\geq\frac{c}{0.5})=1-\Phi(2c).$$

$$\Rightarrow \Phi(2c)=0.975\Rightarrow 2c=1.96\Rightarrow c=0.98.$$

(2)上述检验问题的接受域为$\{|\overline{X}-6|<0.98\}$. 要求$\beta:=P_{\mu=6.5}\{|\overline{X}-6|<0.98\}$.

当$\mu=6.5$时，$\overline{X}\sim N(6.5,\frac{2^2}{16})$,即$\overline{X}\sim N(6.5,0.5^2)$.

$P_{\mu=6.5}\{|\overline{X}-6|<0.98\}=P_{\mu=6.5}\{-0.98<\overline{X}-6<0.98\}$ (三边减0.5，后除0.5)

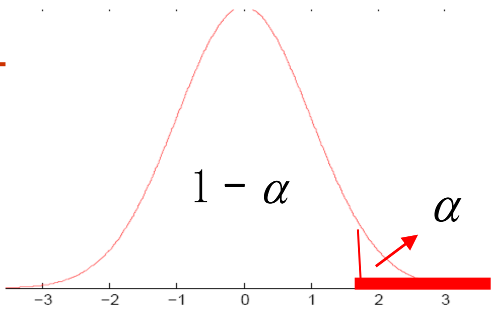
$$=P_{\mu=6.5}\{\frac{-1.48}{0.5}<\frac{\overline{X}-6.5}{0.5}<\frac{0.48}{0.5}\}=\Phi(0.96)-\Phi(-2.96)$$

$$=\Phi(0.96)+\Phi(2.96)-1=0.83.$$

二、检验的p值

- **Why need the p-value?**
  - How do we choose significant level α ?
- If it seems too dangerous to reject true H_0, we choose a small α. How small is small? 0.01 or 001?
  - RR method produces yes or no response, slightly different *α* could lead to different conclusion.
  - How to take full advantage of the information available from the test result?
  - We need probability value or p-value.

回顾“牛奶掺水”问题



$H_0: \mu \le \mu_0 = -0.545$ (即设牛奶未掺水);

$H_1: \mu > \mu_0$ (即设牛奶已掺水).

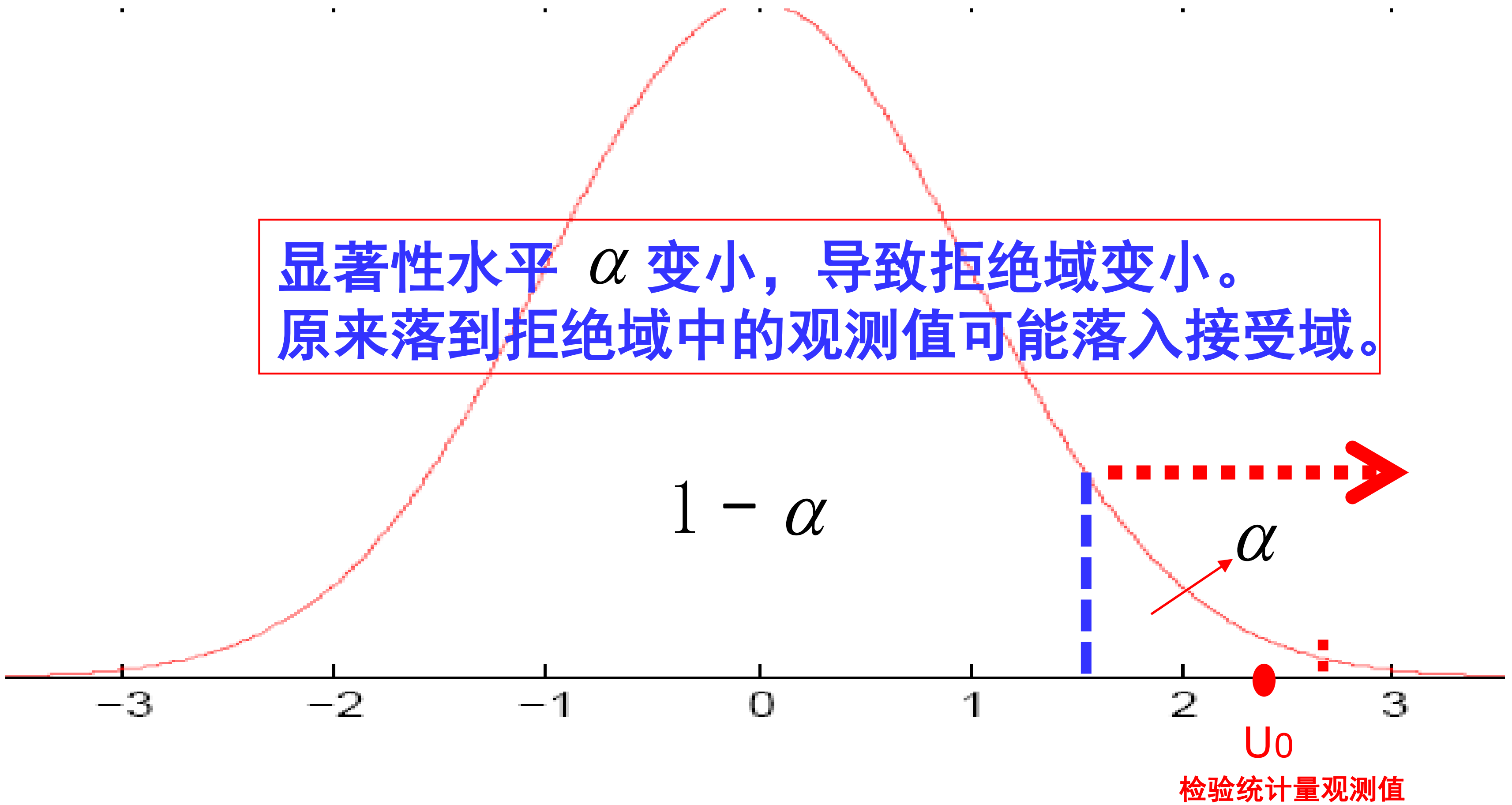
检验统计量：$U = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$，　拒绝域：$\mathbf{W} = \{U \mid U \ge u_{1-\alpha}\}$

(由观测值计算得到的U值为2.80)

显著性水平α	**0.05**	**0.01**	**0.001**
拒绝域	**$U \ge 1.645$**	**$U \ge 2.33$**	**$U \ge 3.01$**
对应的结论	拒绝H_0	拒绝H_0	接受H_0

在较大的显著性水平下得到拒绝原假设的结论，
可能在较小的显著性水平下得到相反的结论。为什么？

不同的显著性水平，可能有不同的结论



不同的显著性水平，可能有不同的结论

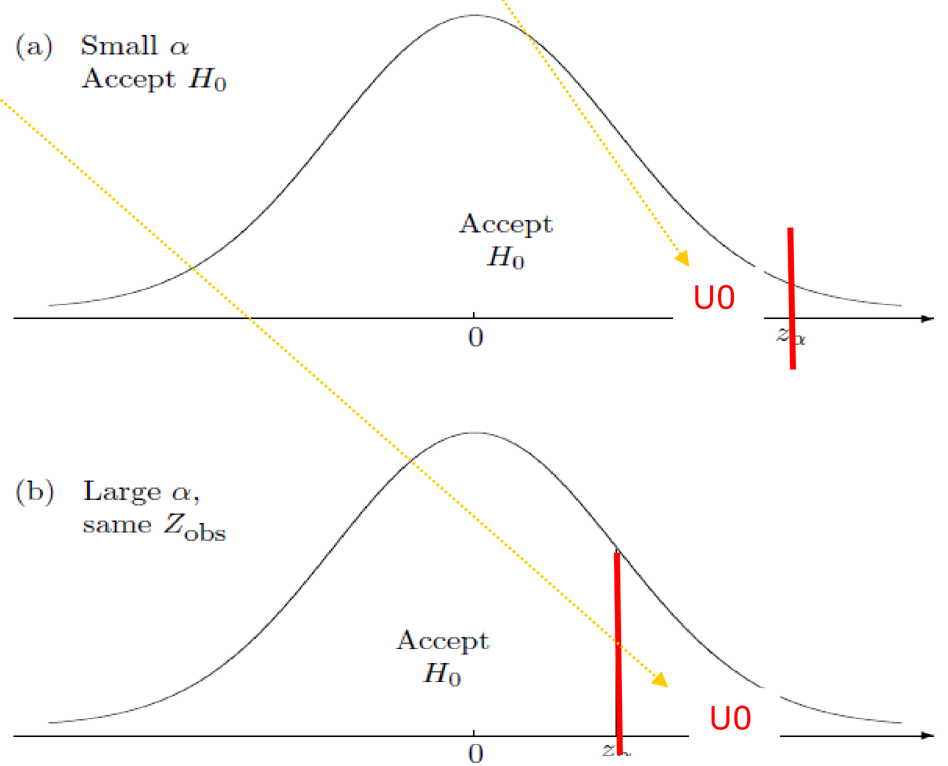
显著性水平$\alpha = 0.05$ 和 $\alpha = 0.001$，可能导致不同结果：

前者的结论是拒绝H_0，而后者的结论是接受H_0.

显著性水平的细小差异，可能导致不同结果。

如果检验结果非常重要，可以依赖于这种结果吗？

利用 p 值。

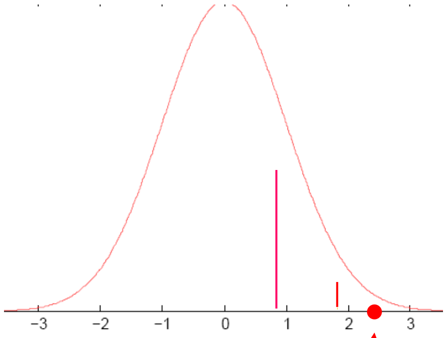


回顾“牛奶掺水”问题

$H_0: \mu \le \mu_0 = -0.545$ (即设牛奶未掺水);

$H_1: \mu > \mu_0$ (即设牛奶已掺水).

检验统计量：$U = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$，　拒绝域：$\mathbf{W} = \{U \mid U \ge u_{1-\alpha}\}$



(由观测值计算得到的U值为2.80)

2.80

对应的尾部面积

显著性水平α	0.05	0.01	?	0.001
拒绝域	$U \ge 1.645$	$U \ge 2.33$	$U \ge 2.8$	$U \ge 3.01$
对应的结论	拒绝H_0	拒绝H_0	拒绝H_0	接受H_0

两类显著性水平之间存在边界：拒绝H_0的最小的显著性水平。
它对应于概率 P(U >= 2.8)

显著性水平比 ? 小，则接受H0；显著性水平比 ? 大，则拒绝H0.

回顾“牛奶掺水”问题

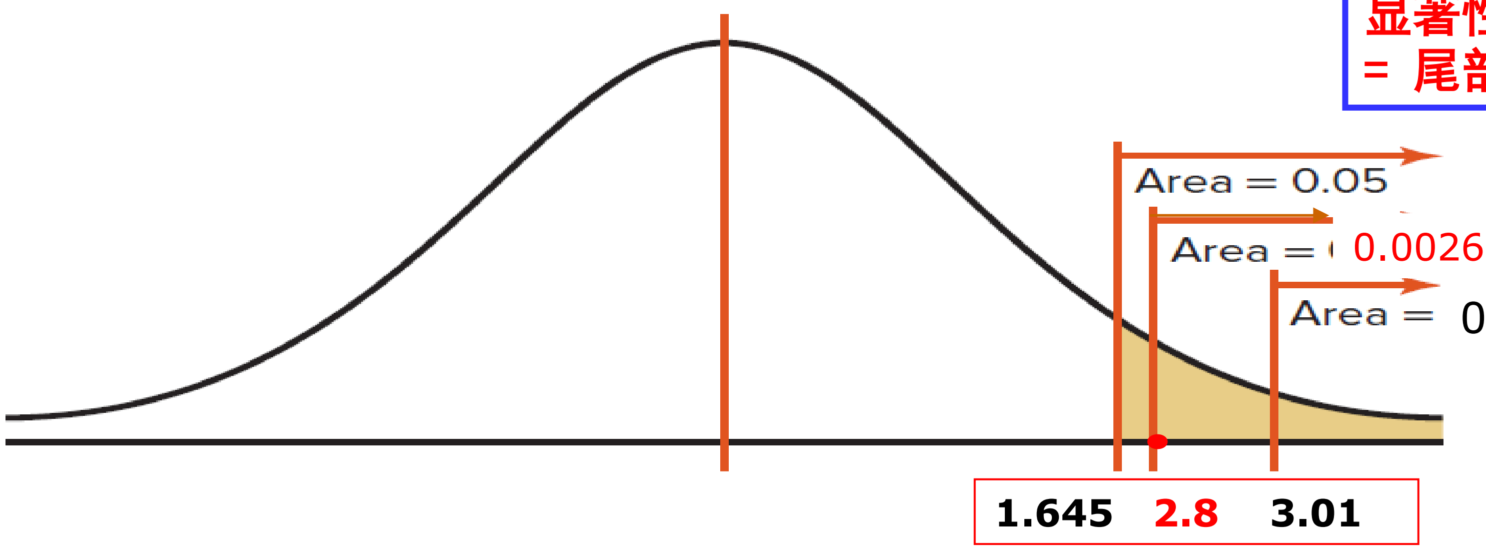
由观测值算得检验统计量为 $U_0 = 2.8$. 问题:

1. 利用该样本观测值能够作出拒绝原假设的 **最小的显著性水平是多少?**

等价

2. 在原假设H0为真的条件下，得到检验统计量为 **当前的观测值或“更极端”的情况的概率是多少?** (更极端指更不利于H0，或更有利于H1)。

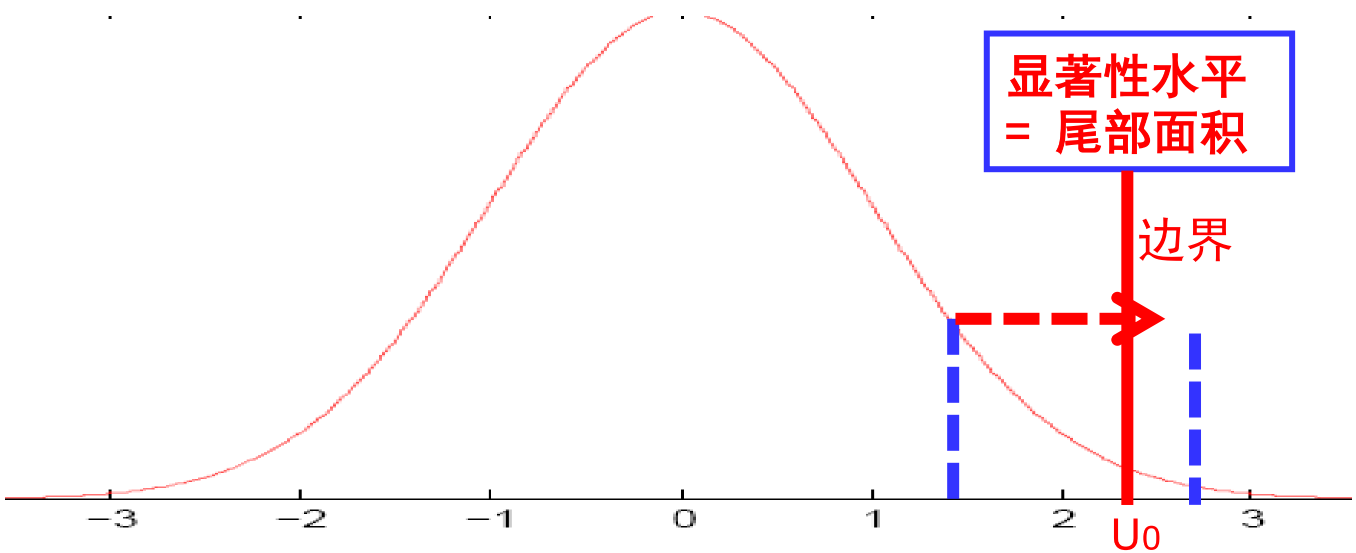
显著性水平 = 尾部面积



定义：p值

在一个假设检验问题中，利用样本观测值能够作出拒绝原假设的最小的显著性水平称为检验的 p 值。

The p-value is the lowest significant level on which we can reject the H0 for the observed data.



检验统计量，固定

回顾“牛奶掺水”问题

对上例：拒绝原假设H_0的最小显著性水平为

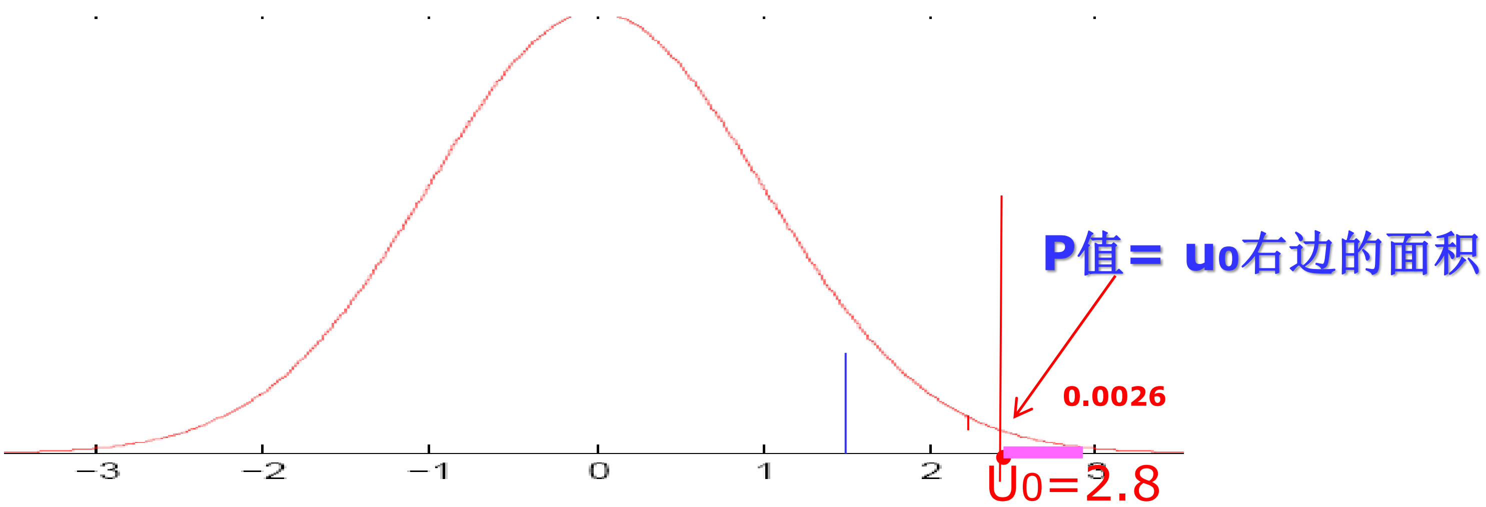
$$p\text{值} = P(U \ge 2.80) = 0.0026.$$

(U0=2.8)

0. 0026是能用观测值u0=2. 80作出“拒绝H_0”的最小显著性水平。这就是p值。

(1) 如果显著性水平$\alpha < p$值，则检验统计量落于接受域，接受H_0

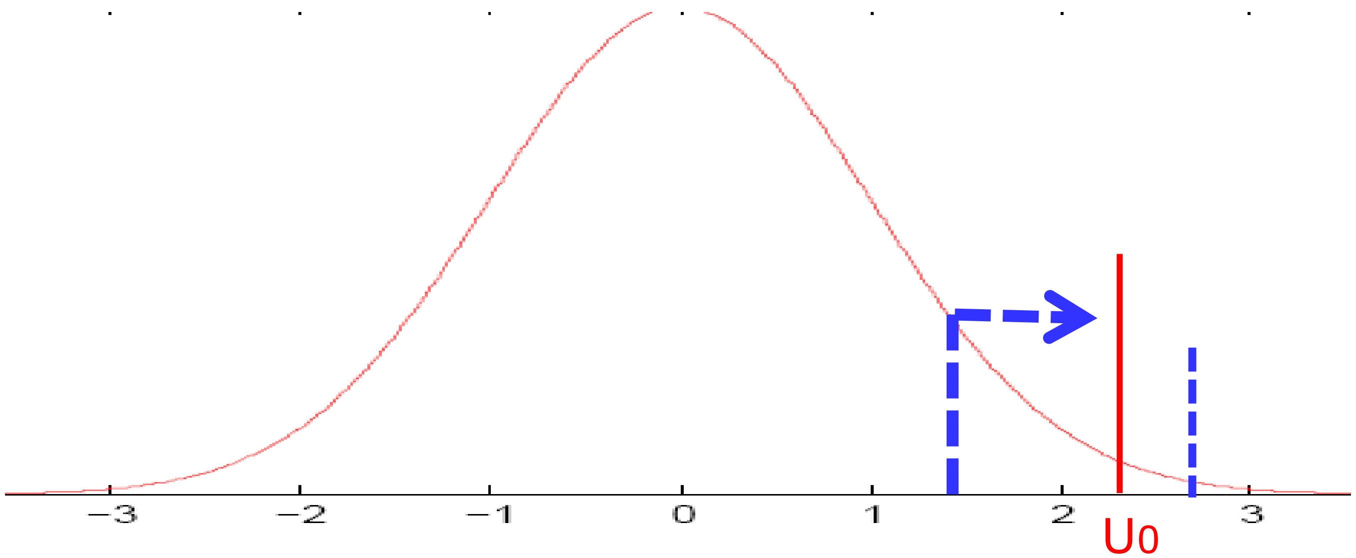
(2) 如果显著性水平$\alpha \ge p$值，则检验统计量落于拒绝域，拒绝H_0



p值的计算

在一个假设检验问题中，利用样本观测值能够作出拒绝原假设的**最小的显著性水平**称为检验的 p 值。

最小显著性水平 = p值 = u_0尾部面积



检验统计量

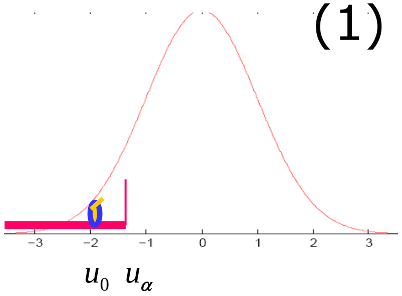
p值的计算：给定检验统计量的观测值 u_0

- **p值是：**在原假设H_0为真的前提下，得到当前观测结果u_0或更极端结果的概率(即尾部面积)。
- 所谓“更极端”，由备择假设决定：
- ✓ 对右侧检验，更大的检验统计量视为更极端；
- ✓ 对左侧检验，更小的检验统计量视为更极端；
- ✓ 对双侧检验，更大或更小的检验统计量都视为更极端。
- **p-value** is the probability of observing test statistics as extreme as or more extreme than the one observed (by assuming H_0 is true).
  It measures how likely we are to get the observed data under H_0.
- **p-value** = tail area (one or two tails) beyond the observed value u_0 of test statistics.

p值的计算：在H0下，得到当前观测结果或更极端结果的概率

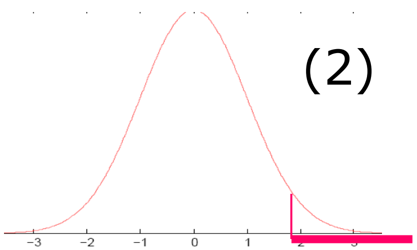
U检验法：检验统计量$U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$，

由观测值计算得到的检验统计量的值 u_0.

(1) 

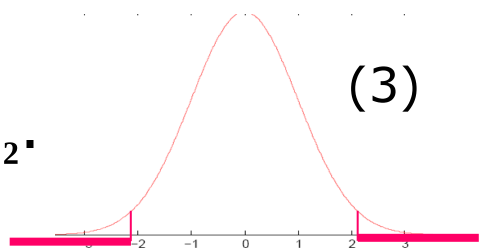
(1)$H_0:\mu\ge\mu_0$ vs $H_1:\mu<\mu_0$；拒绝域：$U\le u_\alpha$.

p值 $=P_{\mu=\mu_0}(U\le u_0)=\Phi(u_0)=u_0$左侧尾部面积。

(2) 

(2)$H_0:\mu\le\mu_0$ vs $H_1:\mu>\mu_0$；拒绝域：$U\ge u_{1-\alpha}$.

p值 $=P_{\mu=\mu_0}(U\ge u_0)=1-\Phi(u_0)=u_0$右侧尾部面积。

(3) 

(3)$H_0:\mu=\mu_0$ vs $H_1:\mu\ne\mu_0$；拒绝域：$|U|\ge u_{1-\alpha/2}$.

p值 $=P_{\mu=\mu_0}(|U|\ge|u_0|)=2(1-\Phi(|u_0|))$

p值的意义：p值的大小意味着什么？

设检验统计量观测值为u_0. 如果H_0为真，观测得到如此结果的可能性有多大？或观测结果是否与H_0相容?

- 小的P值：在H_0为真时, 出现如此结果不太可能。然而，确实观测到如此结果。从而数据与H_0不相容，拒绝H_0.
- 大的P值：在H_0为真时, 出现如此结果（甚至更极端）是非常可能的。从而没有观测到与H_0的矛盾，不拒绝H_0.

- If p-value is small, we reject H_0.
- If p-value is not small, we cannot reject H_0.
- p-value indicates strength of evidence against H_0.
- The smaller the p-value, the stronger evidence against H_0.

p值的意义：p值的大小意味着什么？

The p-value describes how surprising our data are when H_0 is true.

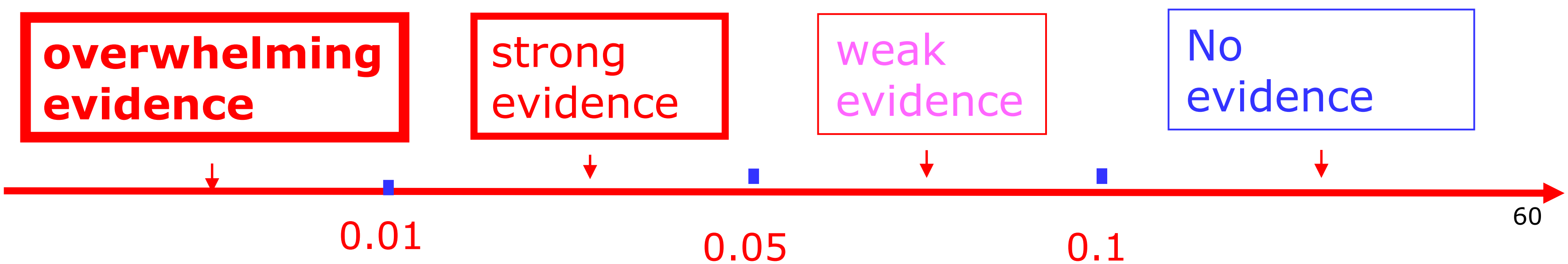
- If p-value < α, then the data is considered to be "rare (or surprising) enough" when H_0 is true. The data provide significant evidence against H_0, thus we reject H_0 and accept H_1. A small p-value provides some *statistical* proof for H1.
- If p-value > α, then our data are not considered to be "surprising enough" when H_0 is true. Our data do not provide enough evidence to reject H_0.

(Logically big p-value does not prove H0;
rather it only shows the lack of evidence against H0).

p值的意义：p值的大小意味着什么？

How small should be the p-value before we decide to reject H_0 ?

- If $p < 0.01$: overwhelming evidence against H_0.
- If $0.01 < p < 0.05$: strong evidence against H_0.
- If $0.05 < p < 0.1$: weak evidence against H_0.
  H_0 is usually not rejected.
- If $p > 0.1$: no evidence against H_0 (H_0 is not rejected.)



p值的意义：p值的大小意味着什么？

P值表示反对原假设 H0 的依据的强弱。

- P值越小，反对 H0 的依据越强，越充分。

  如果P值很小,说明H0为真时，这种情况发生的概率很小.

  而如果出现了，根据小概率原理，我们就有理由拒绝H0.

- P值很小，表示可以在很小的显著性水平下拒绝 H0.

  (即在很小的犯第一类错误的概率的情况下拒绝 H0.)

Steps of Hypothesis Testing

Procedure Table

Solving Hypothesis-Testing Problems (*P*-Value Method)

Step 1 State the hypotheses and identify the claim.

Step 2 Compute the test value.

Step 3 Find the *P*-value.

Step 4 Make the decision.

Step 5 Summarize the results.

Decision Rule When Using a *P*-Value

If *P*-value $\leq \alpha$, reject the null hypothesis.

If *P*-value $> \alpha$, do not reject the null hypothesis.

Testing H_0 with a P-value

For	$\alpha < P$,	accept H_0
For	$\alpha > P$,	reject H_0

Practically,

If	$P < 0.01$,	reject H_0
If	$P > 0.1$,	accept H_0

Rejection region vs p-value approach

- Given significance level a, two approaches always arrive at the same conclusion.
- If *p*-value <= a, then the test statistic falls into RR and vice versa.
- The *p*-value provides additional information
  - Decision for every possible a.
  - Strength of the rejection of H_0.
- RR provides a decision only for specified a.

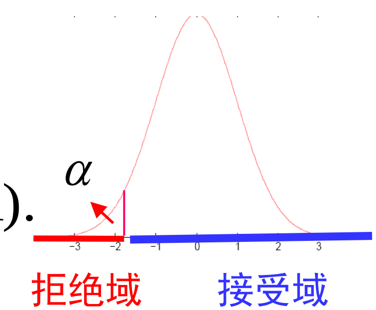
三、关于均值的检验（方差未知）

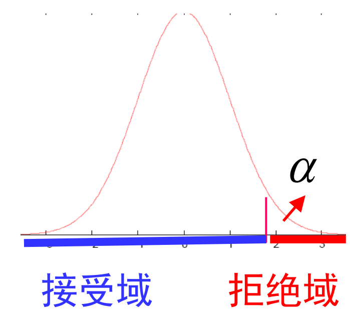
t 检验法：

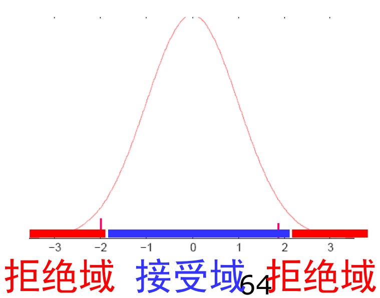
$(1) H_0: \mu \geq \mu_0$　vs　$H_1: \mu < \mu_0$；拒绝域：$t \leq t_\alpha(n-1)$.

$(2) H_0: \mu \leq \mu_0$　vs　$H_1: \mu > \mu_0$；拒绝域：$t \geq t_{1-\alpha}(n-1)$.

$(3) H_0: \mu = \mu_0$　vs　$H_1: \mu \neq \mu_0$；拒绝域：$|t| \geq t_{1-\alpha/2}(n-1)$.







对三种假设检验问题：

* 检验统计量均为：$t = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$；

* 都只要在 $\mu = \mu_0$ 控制犯第I类错误的概率 $= \alpha$ 即可；

* 在 $\mu = \mu_0$ 时，$t = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$.

由 P_{μ_0}(样本落在拒绝域) $= \alpha \Rightarrow$ 可确定临界值。

(拒绝域形式看备择假设)

t检验的p值

t 检验法：检验统计量 $t=\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$,

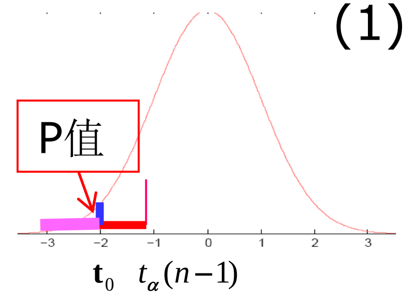
由观测值计算得到的统计量的值 $\mathbf{t}_0$.

(1) $H_0:\mu\ge\mu_0$ vs $H_1:\mu<\mu_0$； 拒绝域: $t\le t_\alpha(n-1)$.

p值 $=P_{\mu=\mu_0}(t\le t_0)=t_0$左侧尾部面积。

当 $t\le t_\alpha(n-1)$ 或 p值 $\le\alpha$ 时，拒绝原假设 H_0.

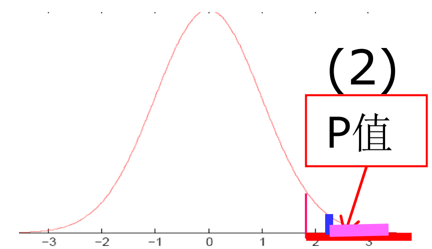
(1)



(2) $H_0:\mu\le\mu_0$ vs $H_1:\mu>\mu_0$； 拒绝域: $t\ge t_{1-\alpha}(n-1)$.

p值 $=P_{\mu=\mu_0}(t\ge t_0)=t_0$右侧尾部面积。

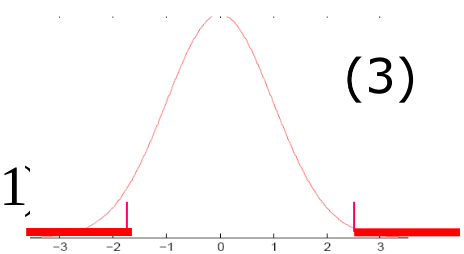
(2)



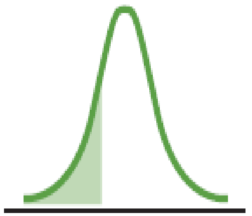
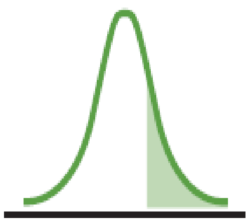
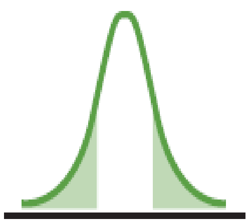
(3) $H_0:\mu=\mu_0$ vs $H_1:\mu\ne\mu_0$； 拒绝域: $|t|\ge t_{1-\alpha/2}(n-1)$

p值 $=P_{\mu=\mu_0}(|t|\ge|t_0|)$

(3)



Alternative hypothesis and p-value area

Statement of H_1		*p*-Value Area	*t*-Curve Region
$\mu < \mu_0$	(less than)	Area to the left of t (even if $t > 0$) $P\{t \le t_{obs}\}$	
$\mu > \mu_0$	(greater than)	Area to the right of t (even if $t < 0$) $P\{t \ge t_{obs}\}$	
$\mu \ne \mu_0$	(not equal)	$2 \times$ area to the right of $\lvert t \rvert$ $P\{\lvert t \rvert \ge \lvert t_{obs} \rvert\}$	

t_{obs} is the observed test statistics

例：某厂生产的铝材的长度服从正态分布，均值设定为$240cm$.

现抽取5件产品，测得样本均值为$\bar{x}=239.5$;样本标准差为$s=0.4$.

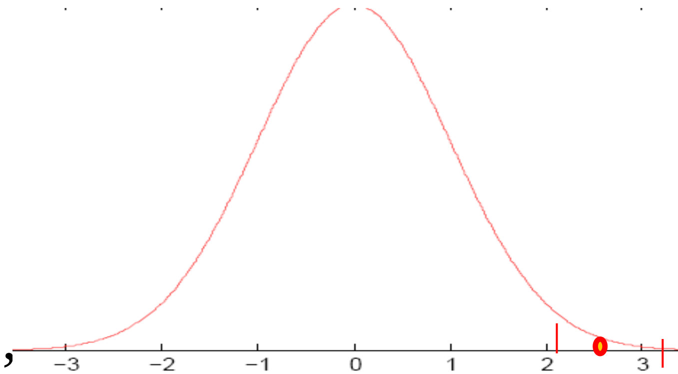
可否认为该厂的产品满足设定要求?(显著性水平为0.05,或**0.01**).

解：需检验假设 $H_0:\mu=240 \quad vs \quad H_1:\mu\neq 240.$

由于总体方差未知，故采用t检验法。

其拒绝域为:$|t|\geq t_{1-\alpha/2}=\begin{cases} t_{0.975}(4)=2.776, & \alpha=0.05. \\ t_{0.995}(4)=4.604, & \alpha=0.01. \end{cases}$

现在,$|t_0|=\left|\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|=\left|\dfrac{239.5-240}{0.4/\sqrt{5}}\right|=2.795.$



*对显著性水平$\alpha=0.05$, $|t_0|=2.795>2.776$,

落在拒绝域中,所以拒绝原假设，即认为长度不满足要求。

*对显著性水平$\alpha=\mathbf{0.01}$, $|t_0|=2.795<4.604$,不落在拒绝域中,

不能拒绝原假设，故接受H_0，认为长度满足要求。

例（续）：

解：需检验假设 $H_0:\mu=240 \quad vs \quad H_1:\mu\neq 240.$

由于总体方差未知，故采用t检验法。

其拒绝域为：$|t|\geq u_{1-\alpha/2}=\begin{cases} t_{0.975}(4)=2.776, & \alpha=0.05. \\ t_{0.995}(4)=4.604, & \alpha=0.01. \end{cases}$

现在，$|t_0|=2.795.$

用 p 值进行检验：

2*(1-pt(2.795, 4))
= 0.04906106

检验的 p值$=P(|t|>2.795)=2P(t>2.795)=0.049$

（---拒绝原假设的最小显著性水平）

对显著性水平$\alpha=0.05>p$值 $\Rightarrow$ 拒绝原假设

对显著性水平$\alpha=0.01<p$值 $\Rightarrow$ 接受原假设

设置

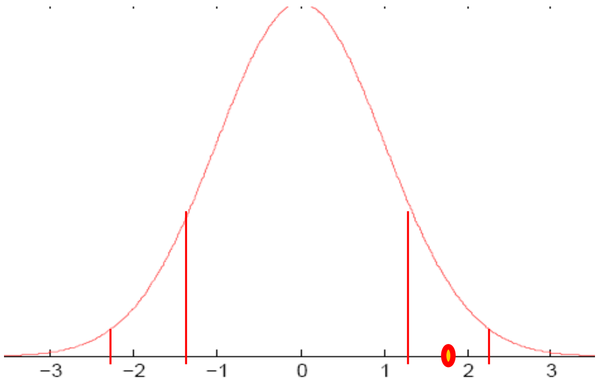
设总体X服从$N(\mu,\sigma^2)$，σ^2未知。$x_1,...,x_n$为来自总体X的样本值。考虑假设检验问题:$H_0:\mu=\mu_0$ vs $H_1:\mu\neq\mu_0$. 若在显著性水平$\alpha=0.05$下拒绝了原假设H_0，则在更小的显著性水平$\alpha=0.01$下，下列结论正确的是

A 必拒绝H0

B 必接受H0

C 犯第一类错误的概率变大

D 可能接受，也可能拒绝H0



提交

例：

设总体$X \sim N(\mu,\sigma^2)$，σ^2未知。$x_1,\ldots,x_n$为来自总体X的样本值。

考虑假设检验问题： $H_0:\mu=\mu_0 \quad vs \quad H_1:\mu\neq\mu_0$.

若在显著性水平$\alpha=0.05$下拒绝了原假设H_0，则在更小的显著性水平$\alpha=0.01$下，下列结论正确的是：

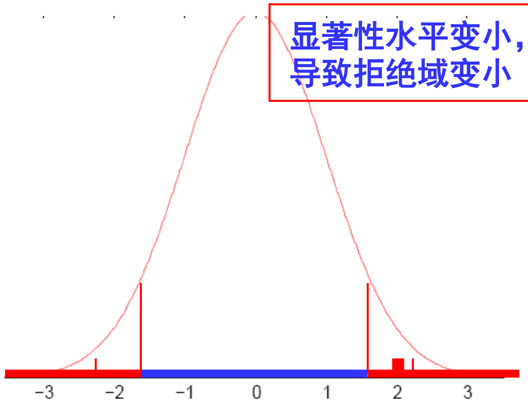
A. 必拒绝H_0 **B.** 必接受H_0

C. 犯第**I**类错误概率变大 **D.** 可能接受，也可能拒绝H_0

解：由于总体方差未知，故采用t检验法。

其拒绝域为：$|t|=\left|\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|\geq t_{1-\alpha/2}$

显著性水平变小时，拒绝域变小。原来落在拒绝域的检验统计量可能落在接受区域（也可能还落在拒绝域）。选**D**。



显著性水平变小，导致拒绝域变小

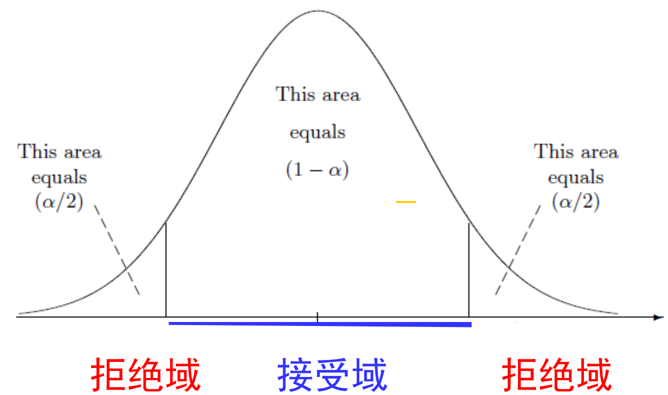
如果显著性水平变大为0.1，结论如何？ → A 必拒绝

四、假设检验与置信区间的关系

考虑正态总体关于均值的检验问题（方差未知）：

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$；拒绝域：$|t| = \left|\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| \geq t_{1-\alpha/2}(n-1).$

接受域W^c：$|t| = \left|\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| < t_{1-\alpha/2}(n-1).$



即 $P_{\mu_0}\left(\left|\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| < t_{1-\alpha/2}(n-1)\right) = 1-\alpha,$

$$\Rightarrow P_{\mu_0}\left(\overline{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1) < \mu_0 < \overline{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right) = 1-\alpha.$$

$\Rightarrow \mu$的置信水平为$1-\alpha$的置信区间：$\overline{X} \pm \dfrac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1).$

四、假设检验与置信区间的关系

反之，若有μ的置信水平为1-α的置信区间：$\overline{X} \pm \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)$,

可得关于 $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

的水平为α的假设检验的拒绝域：$|t| = \left| \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \right| \geq t_{1-\alpha/2}(n-1)$.

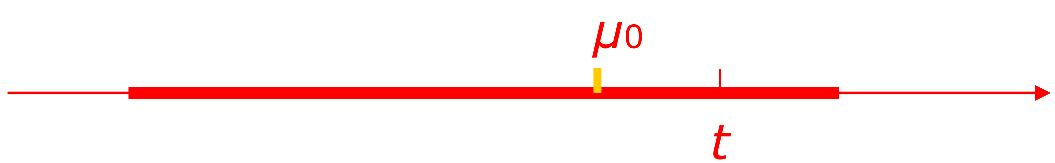
检验统计量落于接受域 等价于 置信区间CI覆盖μ0

对单侧问题，类似。

CI: another way to conduct a test

“检验统计量落于接受域 ” 等价于“置信区间CI覆盖 μ_0 ”.

- We can convert a 1-α CI into an α level test, and vice versa.
- H_0 is not rejected at α level, if the null value is covered by (1- α)100% CI.
- H_0 is rejected if CI does not cover the null value.
- μ_0 lies in CI if and only if the test accepts.
- CI consists precisely of all those values of μ_0 for which H_0: $\mu = \mu_0$ is not rejected.



五、关于正态总体方差的检验

设$x_1,\dots,x_n$是来自正态总体$N(\mu,\sigma^2)$, μ,σ^2未知.

*关于方差有以下三种假设检验:

(1). $H_0:\sigma^2 \geq \sigma_0^2$ **vs** $H_1:\sigma^2 < \sigma_0^2$;

(2). $H_0:\sigma^2 \leq \sigma_0^2$ **vs** $H_1:\sigma^2 > \sigma_0^2$;

(3). $H_0:\sigma^2 = \sigma_0^2$ **vs** $H_1:\sigma^2 \neq \sigma_0^2$;

考虑：

(2). $H_0:\boldsymbol{\sigma}^2 \le \boldsymbol{\sigma}_0^2$ **vs** $H_1:\boldsymbol{\sigma}^2 > \boldsymbol{\sigma}_0^2$;

样本方差S^2为总体方差σ^2的无偏估计。

当H_1为真时，样本方差的观测值往往偏大。

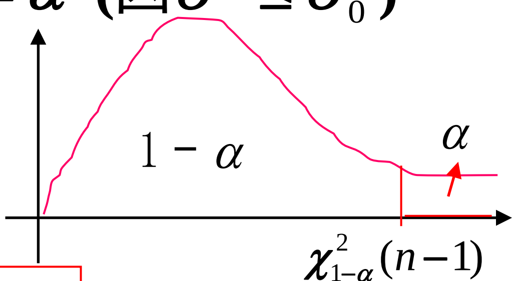
检验统计量为：$\chi^2 = \dfrac{(n-1)S^2}{\boldsymbol{\sigma}_0^2}$；拒绝域的形式：$W=\{\chi^2 \ge c\}$.

希望 **P{犯第I类错误}** $=$ **P{当**H_0为真时拒绝$H_0\} \le \boldsymbol{\alpha}$;

P{当H_0为真时拒绝$H_0\} = \mathbf{P}_{\sigma^2\le\sigma_0^2}\left\{\dfrac{(n-1)S^2}{\boldsymbol{\sigma}_0^2} \ge c\right\}$

$$= \mathbf{P}_{\sigma^2\le\sigma_0^2}\left\{\frac{(n-1)S^2}{\boldsymbol{\sigma}^2} \ge c\,\frac{\boldsymbol{\sigma}_0^2}{\boldsymbol{\sigma}^2}\right\} \le \mathbf{P}_{\sigma^2\le\sigma_0^2}\left\{\frac{(n-1)S^2}{\boldsymbol{\sigma}^2} \ge c\right\} = \boldsymbol{\alpha}\ \ (\text{因}\,\boldsymbol{\sigma}^2 \le \boldsymbol{\sigma}_0^2)$$

因 $\chi^2 = \dfrac{(n-1)S^2}{\boldsymbol{\sigma}^2} \sim \chi^2(n-1) \quad \Rightarrow \quad c = \chi^2_{1-\alpha}(n-1).$



这个概率是sigma 的增函数

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

考虑：

(3). $H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2\neq\sigma_0^2$;

注意到S^2是σ^2的无偏估计，

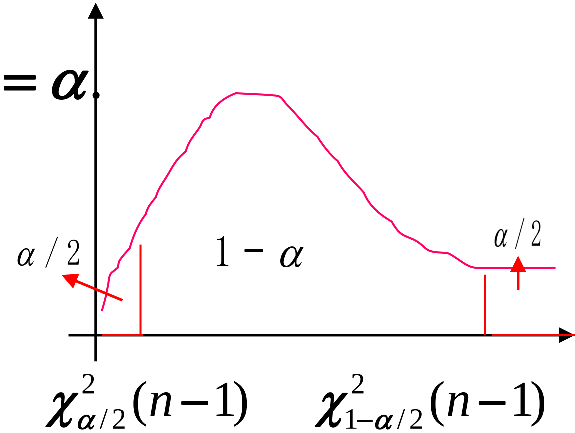
当H_0为真时，S^2/σ_0^2一般应在1附近；而不应过分大或过分小。

检验统计量为：$\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}$;

拒绝域的形式：$W=\{\chi^2\le c_1\}\cup\{\chi^2\ge c_2\}$.

希望 P{犯第I类错误} = P{当H_0为真时拒绝H_0} = α;

即：$P_{\sigma^2=\sigma_0^2}\left\{\dfrac{(n-1)S^2}{\sigma_0^2}\le c_1\right\}+P_{\sigma^2=\sigma_0^2}\left\{\dfrac{(n-1)S^2}{\sigma_0^2}\ge c_2\right\}=\alpha$.



因当H_0为真时 $\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$

$\Rightarrow$ 可取 $c_1=\chi^2_{\alpha/2}(n-1)$, $c_2=\chi^2_{1-\alpha/2}(n-1)$.

χ^2 检验法：

(1). $H_0:\sigma^2 \ge \sigma_0^2$ **vs** $H_1:\sigma^2 < \sigma_0^2$；拒绝域：$\chi^2 \le \chi_\alpha^2(n-1)$.

(2). $H_0:\sigma^2 \le \sigma_0^2$ **vs** $H_1:\sigma^2 > \sigma_0^2$；拒绝域：$\chi^2 \ge \chi_{1-\alpha}^2(n-1)$.

(3). $H_0:\sigma^2 = \sigma_0^2$ **vs** $H_1:\sigma^2 \ne \sigma_0^2$；拒绝域：$\chi^2 \le \chi_{\alpha/2}^2(n-1)$

或$\chi^2 \ge \chi_{1-\alpha/2}^2(n-1)$.

对三种假设检验问题：

* 检验统计量均为：$\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$；(对应样本的值：$\chi_0^2$)

* 都只要在$\sigma^2 = \sigma_0^2$控制犯第I类错误的概率$=\alpha$即可；

* 在$\sigma^2 = \sigma_0^2$时，$\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$.

由$P_{\sigma_0^2}$(样本落在拒绝域)$=\alpha \Rightarrow$可确定临界值。

(1) 如果$\alpha < p$值,则在显著性水平α下，接受H_0

(2) 如果$\alpha \ge p$值,则在显著性水平α下，拒绝H_0

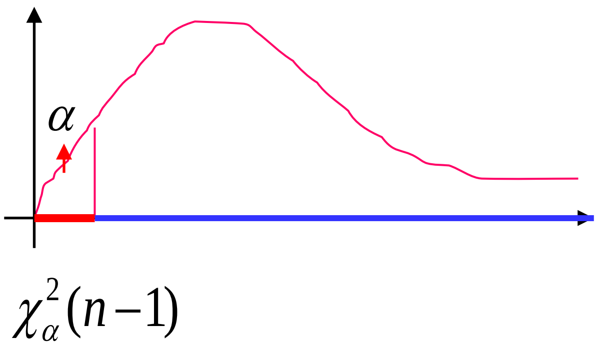
χ^2 检验法：p值

(1). $H_0:\sigma^2 \ge \sigma_0^2$ vs $H_1:\sigma^2 < \sigma_0^2$; 拒绝域:$\chi^2 \le \chi_\alpha^2(n-1)$. **p值 P($\chi^2 \le \chi_0^2$)**

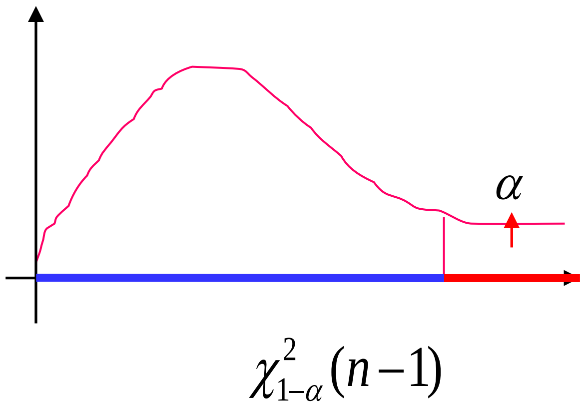
(2). $H_0:\sigma^2 \le \sigma_0^2$ vs $H_1:\sigma^2 > \sigma_0^2$; 拒绝域:$\chi^2 \ge \chi_{1-\alpha}^2(n-1)$. **p值 P($\chi^2 \ge \chi_0^2$)**

(3). $H_0:\sigma^2 = \sigma_0^2$ vs $H_1:\sigma^2 \ne \sigma_0^2$; 拒绝域:$\chi^2 \le \chi_{\alpha/2}^2(n-1)$ 或 $\chi^2 \ge \chi_{1-\alpha/2}^2(n-1)$

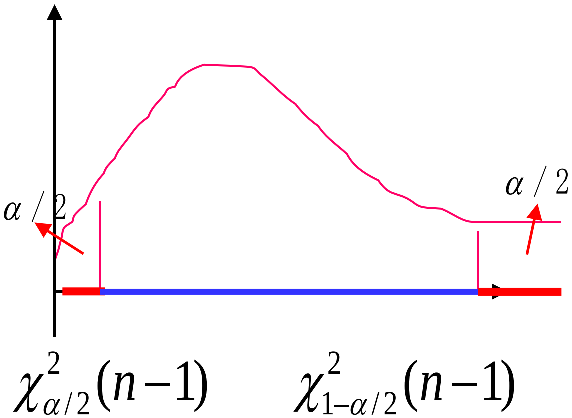
p值 2min{P($\chi^2 \le \chi_0^2$),P($\chi^2 \ge \chi_0^2$)}



(1)



(2)



(3)

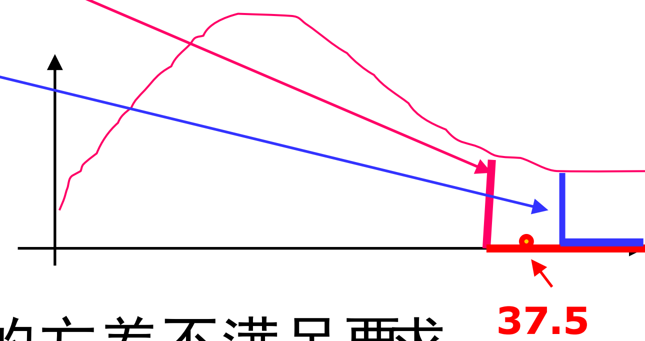
例：

某厂生产的钢板重量服从正态分布，设定重量的方差不得超过0.016. 现抽取25件产品，测得样本方差$s^2=0.025$. 问可否认为该厂生产的钢板重量的方差满足设定要求?(显著性水平为0.05,或0.01).

解：需检验假设 $H_0:\sigma^2\le 0.016 \quad vs \quad H_1:\sigma^2>0.016$. 采用$\chi^2$检验法。

其拒绝域为：$\chi^2\ge\chi^2_{1-\alpha}=\begin{cases}\chi^2_{1-0.05}(24)=36.415, & \alpha=0.05.\\ \chi^2_{1-0.01}(24)=42.980, & \alpha=0.01.\end{cases}$

现在，$\chi^2=\dfrac{(n-1)s^2}{\sigma_0^2}=\dfrac{24\times 0.025}{0.016}=37.5.$



$*$对显著性水平$\alpha=0.05$, $\chi^2=37.5>36.415$,

落在拒绝域中,所以拒绝原假设，即认为重量的方差不满足要求。

$*$对显著性水平$\alpha=0.01$, $\chi^2=37.5<42.980$,

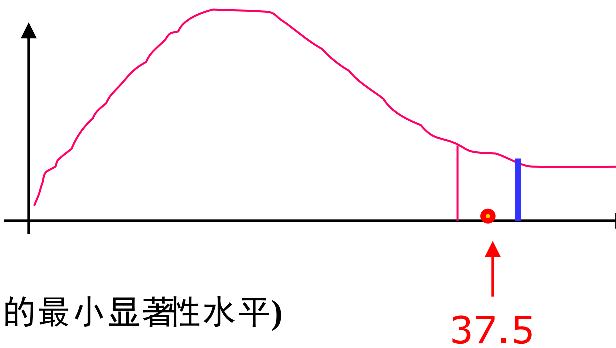
不落在拒绝域中，不能拒绝原假设，故接受H_0，认为满足要求(更慎重)

例（续）

解：需检验假设 $H_0: \sigma^2 \le 0.016 \quad vs \quad H_1: \sigma^2 > 0.016$. 采用$\chi^2$检验法。

其拒绝域为：$\chi^2 \ge \chi^2_{1-\alpha} = \begin{cases} \chi^2_{1-0.05}(24) = 36.415, & \alpha = 0.05. \\ \chi^2_{1-0.01}(24) = 42.980, & \alpha = 0.01. \end{cases}$

现在，$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{24 \times 0.025}{0.016} = 37.5.$



检验的 p 值：$\mathbf{p = P(\chi^2 \ge 37.5) = 0.039.}$ (拒绝原假设的最小显著性水平)

1-pchisq(37.5, 24)= 0.0389818

对显著性水平$\alpha = 0.05$, $\alpha > p \Rightarrow$ 拒绝原假设

对显著性水平$\alpha = 0.01$, $\alpha < p \Rightarrow$ 接受原假设

(1) 如果$\alpha < p$值,则在显著性水平α下，接受H_0

(2) 如果$\alpha \ge p$值,则在显著性水平α下，拒绝H_0

假设检验总结

- **建立假设：**
  - 单侧检验 vs 双侧检验
  - 均值检验 vs 方差检验
  - 单 总 体 vs ~~双总体(成对数据？)~~

是否有显著差异？ → 双侧检验
是否有显著提高(改进)？ → H1：“ > ”
是否有显著降低? → H1：“ < ”

- **检验统计量**
  - 均值检验：U检验法（方差已知）vs t检验法（方差未知）
  - 方差检验：χ^2 检验（单总体），~~F检验法（双总体）~~

- **判断：**
  - 拒绝域（临界值）方法
  - p值方法
  - CI方法

The End of Chapter 7

关于两个正态总体的检验不要求。