假设检验

- 什么是假设?(hypothesis)
  - 对总体分布中未知的参数数值所作的一种陈述
    - 总体参数包括**总体均值**、**比例**、**方差**等
    - 检验**之前**必需陈述
    - 比如： $\mu = \mu_0$

      $\mu < \mu_0$

      $\mu > \mu_0$

- 什么是假设检验?(hypothesis testing)
  - 事先对总体参数作出某种假设，然后利用样本信息来判断假设是否成立
  - 按统计方法分，有参数假设检验和非参数假设检验
  - 均采用逻辑上的反证法，依据统计上的小概率原理

原假设与备择假设

- ➡ **什么是原假设？(null hypothesis)**

1. 待检验的假设，又称“0假设”

2. 表示为 H_0
   - H_0：μ = 某一数值
   - 指定为 = 号，或 ≤ 或 ≥
   - 例如, H_0：$\mu = 3190$（克）

- ➡ **什么是备择假设？(alternative hypothesis)**

1. 与原假设对立的假设，也称“研究假设”
2. 表示为 H_1
   - H_1：μ <某一数值，或μ >某一数值
   - 例如, H_1：μ < 3910(克)，或μ >3910(克)

检验中的两类错误
（对H_0的决策风险）

- 对H_0的决策风险来源于样本的随机性

1. **第一类错误（弃真错误）**
   - 原假设为真时拒绝原假设
   - 会产生一系列后果
   - 第一类错误的错判概率为α
     - 被称为显著性水平
2. **第二类错误（采伪错误）**
   - 原假设为假时接受原假设
   - 第二类错误的错判概率为β

	接受H_0	拒绝H_0
H_0为真	正确决策	第一类错误
H_0为假	第二类错误	正确决策

两类错误的控制原则

- 统计学家Neyman和Pearson提出的原则
  - 在控制α的情况下，使β尽量小。
- 控制原则含义的解读
  - H_0要受到保护，使它不至于轻易被否定。
  - 若检验结果否定了H_0，则说明样本信息给出的否定理由充分；同时，α受到控制，亦即作出否定H_0的可靠程度（$1-\alpha$）得到了保证。
- 通常α值控制在1%~5%之间，β值控制在10%~30%之间。

显著性水平α (significant level)

- ➡ 什么是显著性水平？

1. 是一个概率值，表示为 α
2. 原假设为真时，拒绝原假设的概率
   - 是对H_0做决策犯一类错误的误判概率
   - 小概率的界限值
   - 它对应的是H_0的拒绝域
   - 由研究者事先确定
3. 显著性水平0.05的意思是：在零假设正确的情况下进行100次抽样，会有5次错误地拒绝了零假设。

检验功效

- 由于β表示为接受不真实原假设结论的概率，那么1-β 就指拒绝不真实原假设的概率。
- 若1-β的数值越接近于1，表明不真实的原假设几乎都能够被拒绝。反之，若1-β接近于0，表明犯第二类错误的可能性很大。
- 因此，1-β可以用来衡量假设检验工作的好坏，理论上称之为检验功效。
- 在α给定的条件下，1-β越大，检验的功效越好。

谨慎对待统计检验的结果

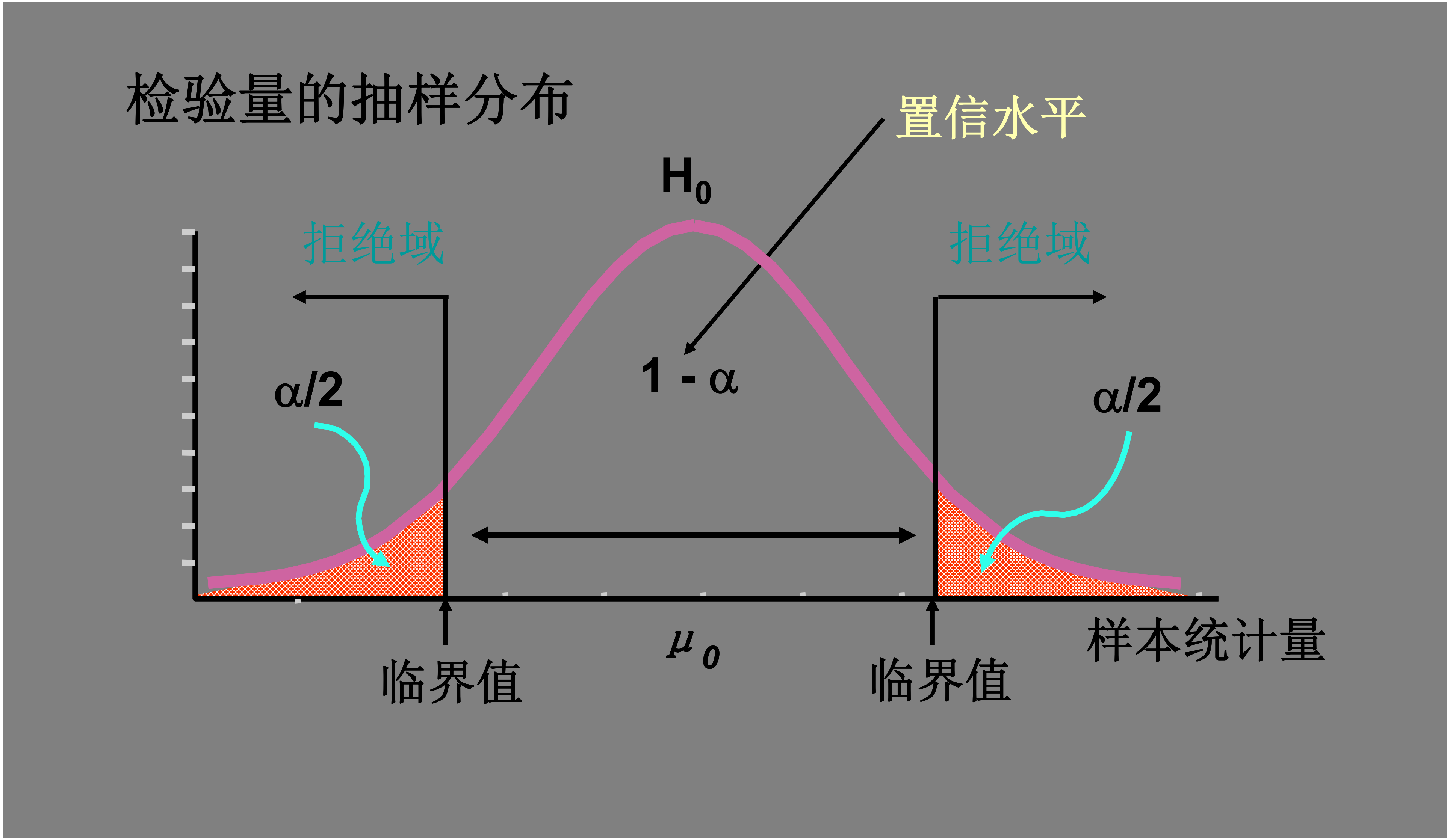
- 在规定的显著性水平α下，拒绝H_0有两种解释：一种是H_0是正确的，观测到的样本数据恰好是不常发生的那一类，另一种是样本数据倒是常见的那一类，只是H_0是错的。
- 如果样本数据没能拒绝H_0，仅仅说明证据不足，无法否定H_0，但这不能说明H_0正确。要想证明H_0正确，则必须知道总体参数。也就是说，我们需要对总体的每个单元进行检验、调查或计数，这通常是不大可行的。代替方法是从总体中抽取一个样本。

H_0的单、双侧检验
（H_0的单、双尾检验）

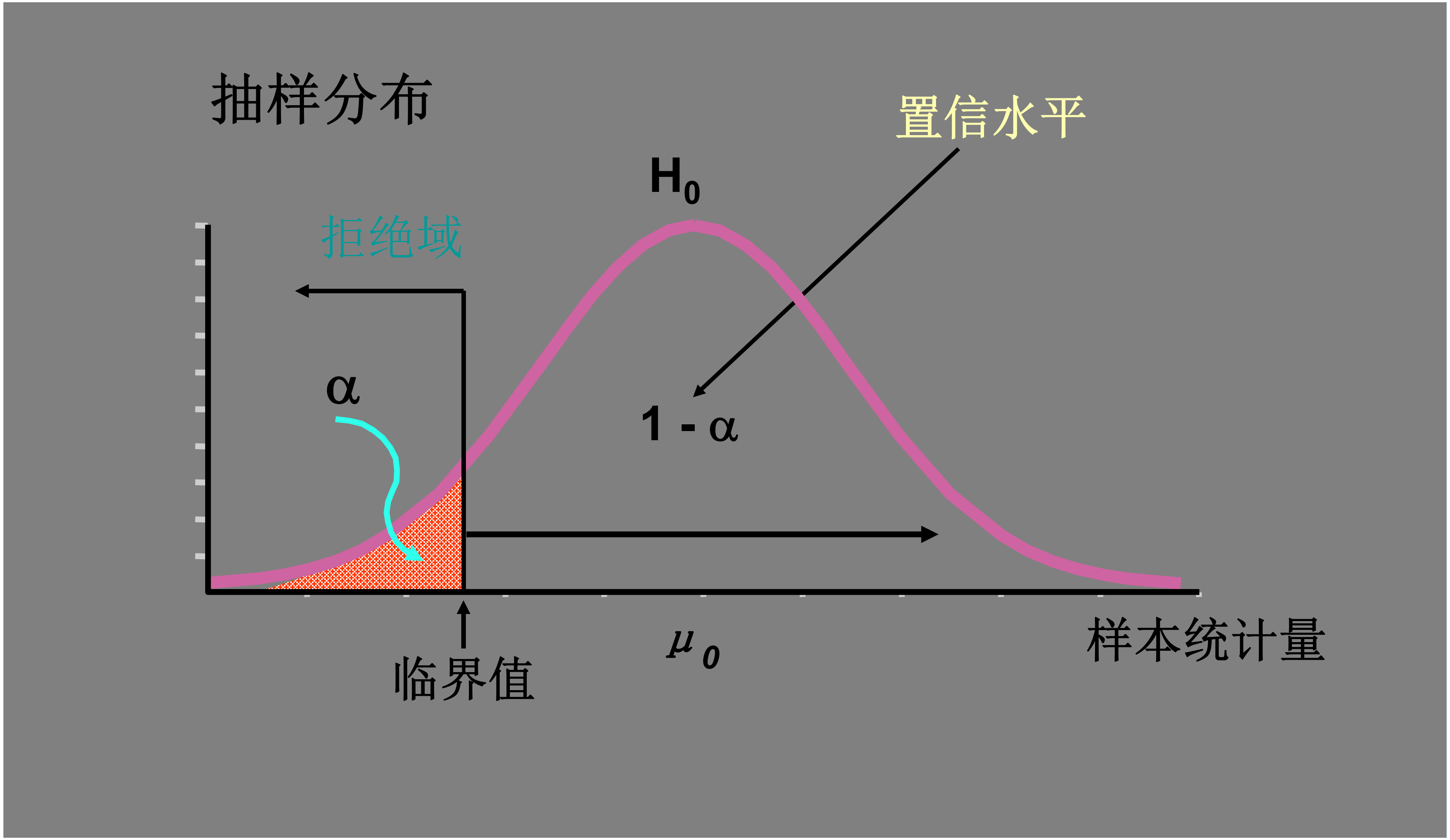
- 由H_1的不同表达形式

假设	**研究的问题**		
	双侧检验	**左侧检验**	**右侧检验**
H_0	$\mu = \mu_0$	$\mu \geq \mu_0$	$\mu \leq \mu_0$
H_1	$\mu \neq \mu_0$	$\mu < \mu_0$	$\mu > \mu_0$

双侧检验图示(显著性水平与拒绝域)



左侧检验图示(显著性水平与拒绝域)



确定适当的检验量

- ➡ 什么是检验统计量？

1. 用于对H_0决策的统计量
2. 选择检验量的方法与参数估计相同，需考虑
   - 正态总体还是非正态总体
   - 是大样本还是小样本
   - 总体方差已知还是未知
3. 检验统计量的基本形式为 $Z=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

对H_0作出统计决策（临界值检验法）

1. 计算检验的统计量
2. 根据给定的显著性水平α，查表得出相应的临界值z_α或$z_{\alpha/2}$， t_α或$t_{\alpha/2}$
3. 将检验统计量的值与α水平的临界值进行比较
4. 得出拒绝或不拒绝H_0的结论

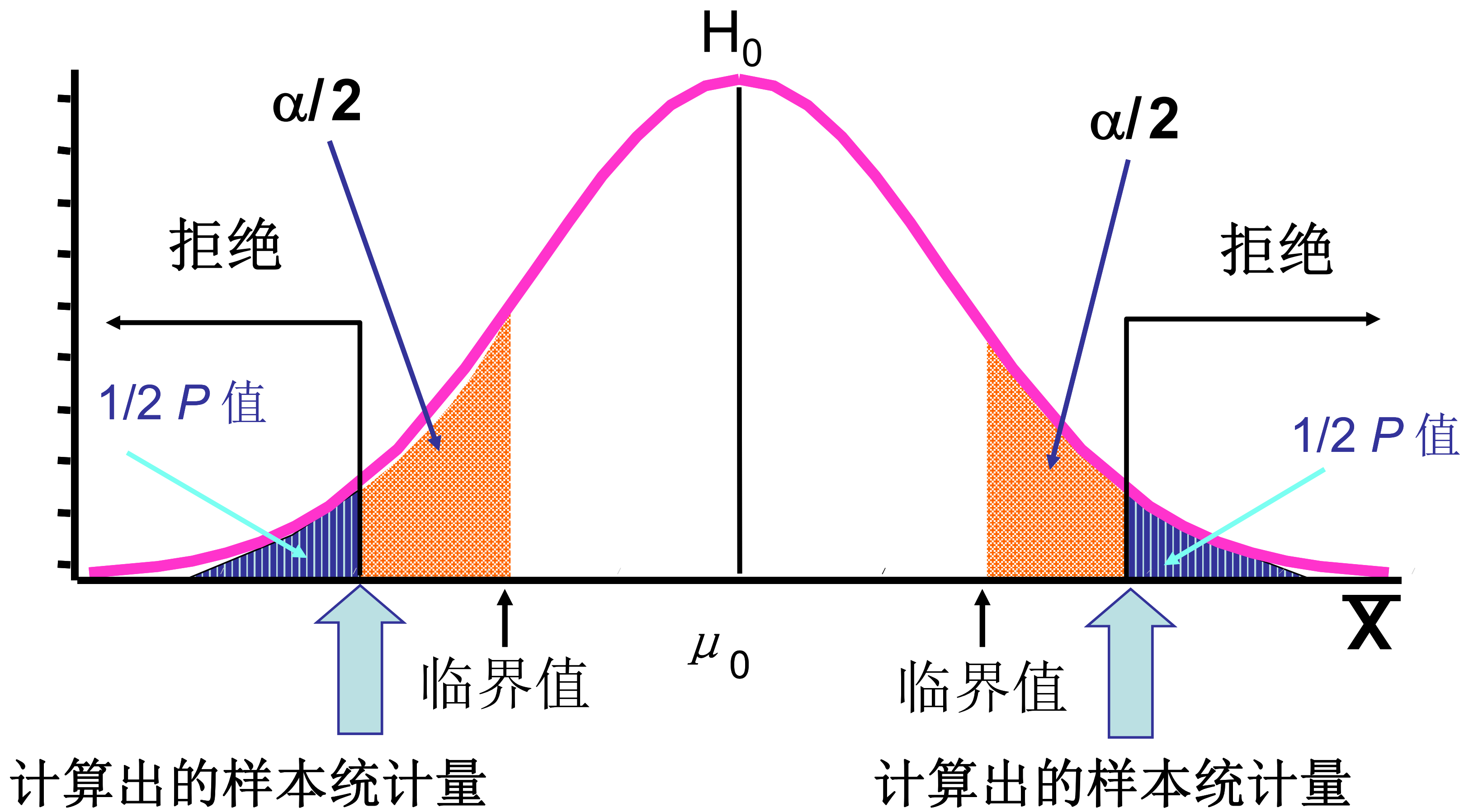
利用P值对H_0进行决策
（P值检验法）

- 什么是*P* 值?(*P*-value)
  - 是一个预拒绝H_0的条件概率值
  - 如果H_0为真，*P*-值是抽样分布中大于或小于计算出的样本统计量的概率
    - 左侧检验时，*P*-值为曲线上方*小于等于*检验统计量部分的面积
    - 右侧检验时，*P*-值为曲线上方*大于等于*检验统计量部分的面积
  - 被称为观察到的(或实测的)显著性水平
    - H_0 能被拒绝的最小值

H_0： $\mu=\mu_0$

H_1： $\mu\neq\mu_0$

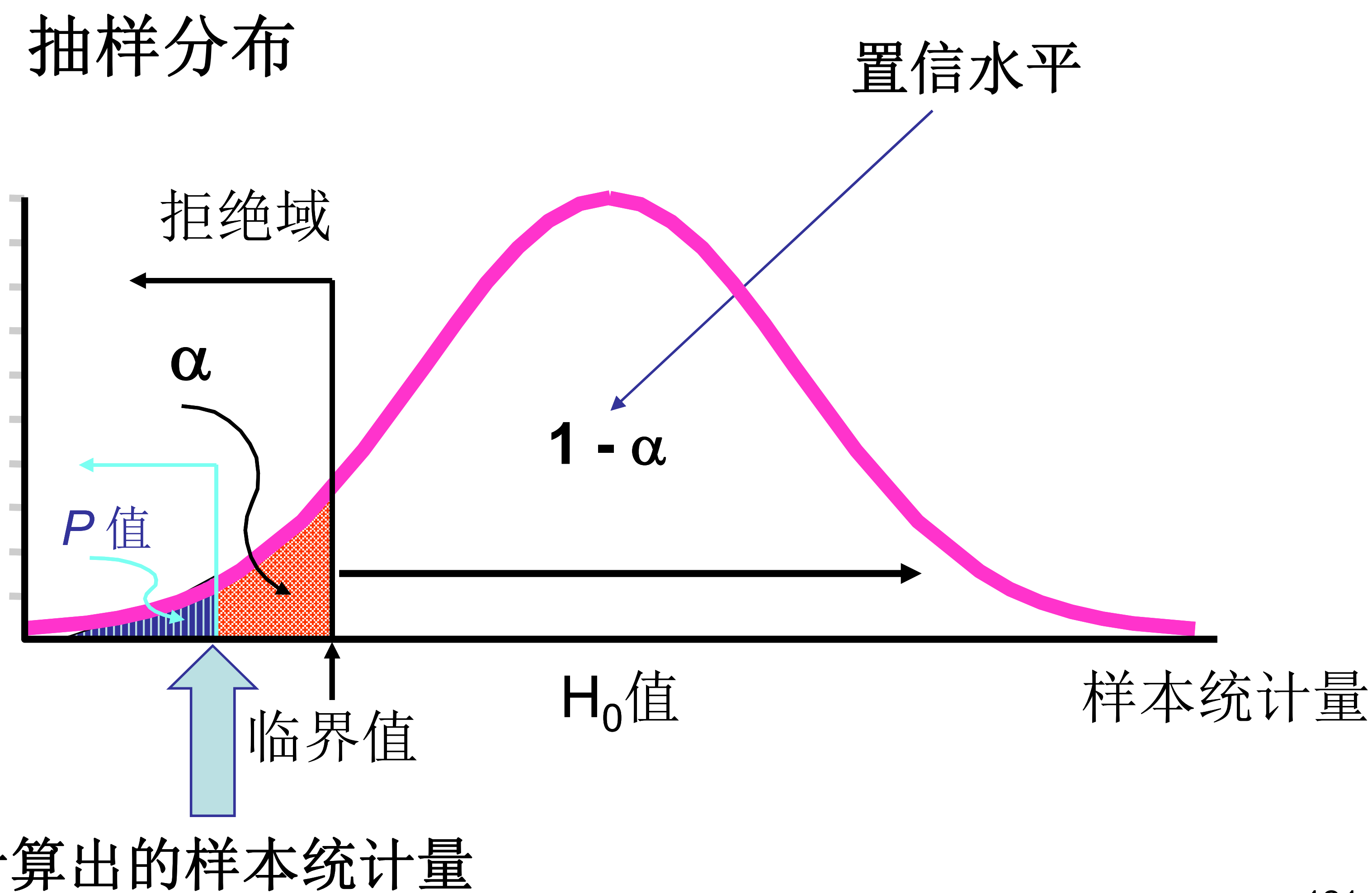
双侧检验的*P* 值



H_0：$\mu \geqslant \mu_0$

H_1：$\mu < \mu_0$

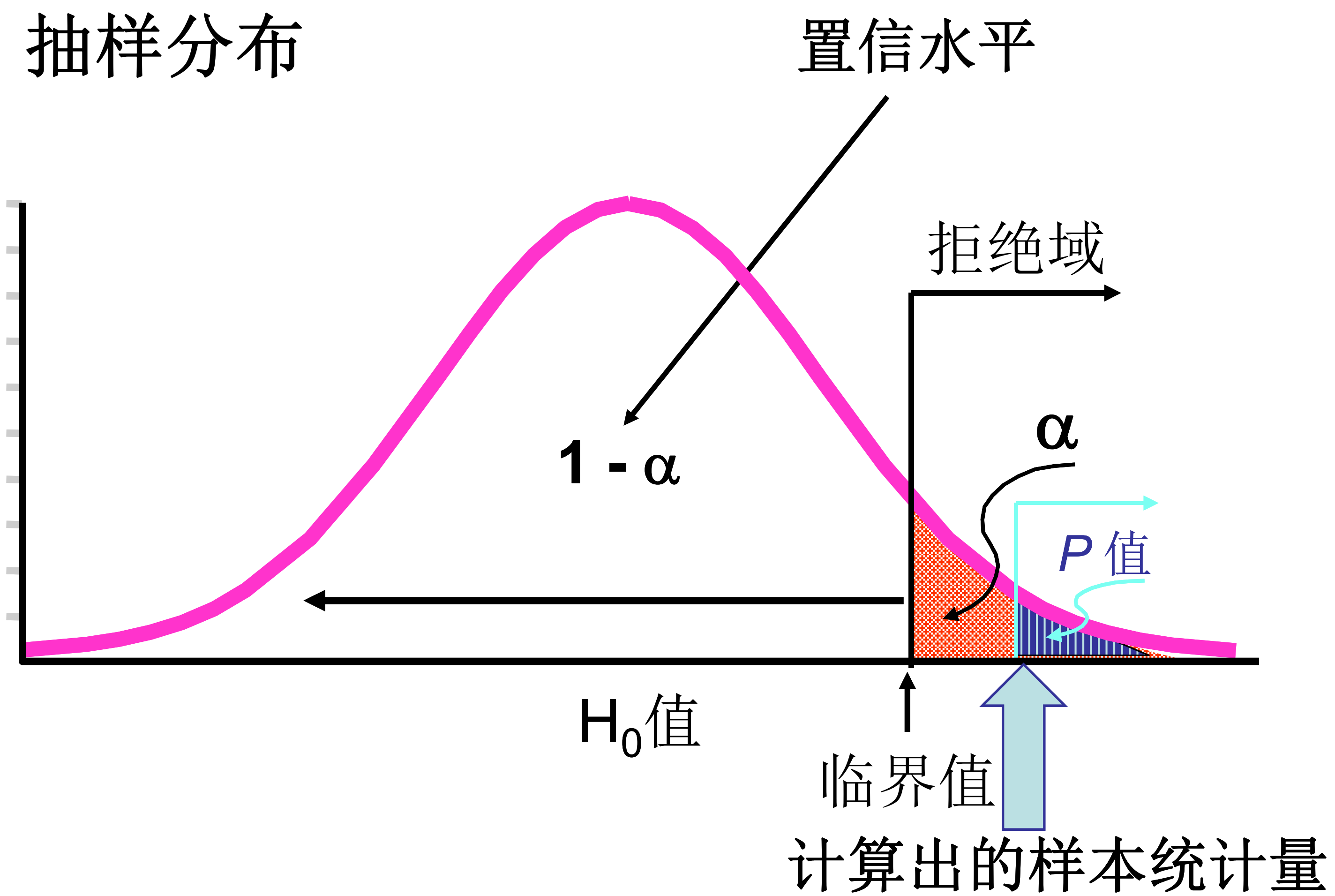
左侧检验的*P* 值



H_0: $\mu \leqslant \mu_0$

H_1: $\mu > \mu_0$

右侧检验的*P* 值



关于*P*-值含义的理解

1. *P*-值反映样本与总体的差别，即度量从样本数据得到的信息对H_0的支持程度
2. *P*-值反映拒绝H_0的力度
3. *P*-值反映拒绝H_0的实际风险（因此*P*-值常常被称为H_0检验的观测显著性水平**(observed level of significance)**）

4. 在统计软件的输出中，通常只输出*P*-值，而由用户去决定*P*-值是多少时拒绝原假设。（因为*P*-值是由数据决定的，显著性水平 a 是由检验者决定的，而不是由计算机给出的。）

5. 在统计软件输出*P*-值的位置，有的用“*P*-value”，有的用significance的缩写“Sig”来标明，也有用概率等符号来表示的。

利用 *P* 值进行检验
(决策准则)

1. 单侧检验
   - 若*P*-值 > α, 不拒绝 H_0
   - 若*P*-值 < α, 拒绝 H_0
2. 双侧检验
   - 若*P*-值 > α/2, 不拒绝 H_0
   - 若*P*-值 < α/2, 拒绝 H_0

临界值检验法的操作步骤

1. 依要回答的问题，建立假设
2. 确定原假设的检验量
3. 确定检验量的抽样分布
4. 根据事先确定的显著性水平α，查表得出相应的临界值z_α或$z_{\alpha/2}$， t_α或$t_{\alpha/2}$
5. 将检验统计量的值与α水平的临界值进行比较
6. 得出拒绝或不拒绝H_0的结论

P-值检验法的操作步骤

1. 依要回答的问题，建立假设
2. 确定原假设的检验量
3. 确定检验量的抽样分布
4. 根据实际情况，事先确定显著性水平α
5. 根据数据计算检验统计量的实现值；
6. 根据这个实现值计算*P*-值；
7. 进行判断：如果*P*-值小于或等于a ，就拒绝零假设，这时犯（第一类）错误的概率最多为a ；如果*P*-值大于a ，就不拒绝零假设，因为证据不足。

假设检验的过程

第1步　　第2步　　第3 步　第4步　　第5步

提出零假设、备择假设 → **确定检验量及其抽样分布** → **选择一个显著性水平α** → **构造H_0的决策规则** → **选取样本，由计算结果做出结论**

→ **不能拒绝H_0**

→ **拒绝H_0，接受H_1**

常用的参数统计检验法

- 检验量的抽样分布是参数统计检验的核心
- 按检验量的抽样分布，最基本的参数检验方法有：

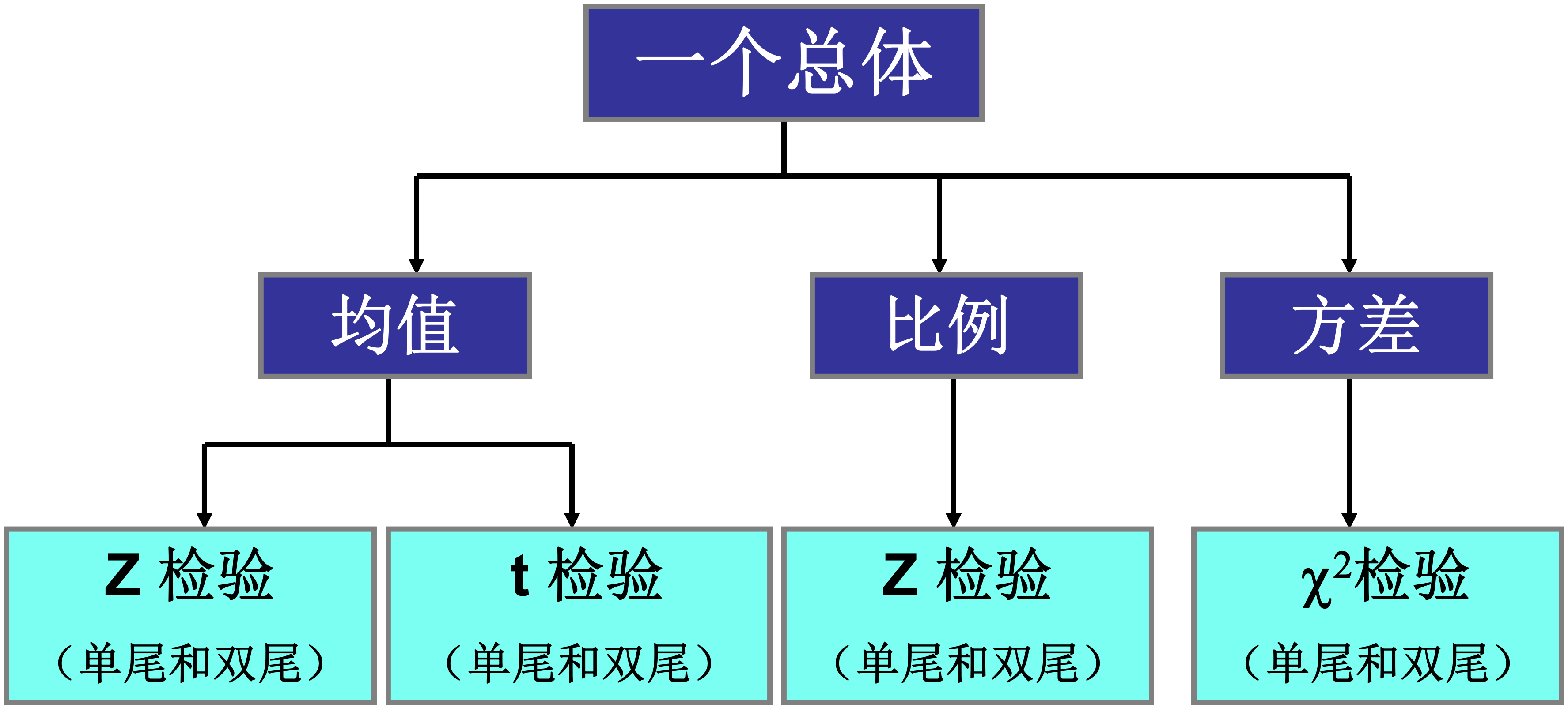
  *Z*检验（或*U*检验）

  t-检验、

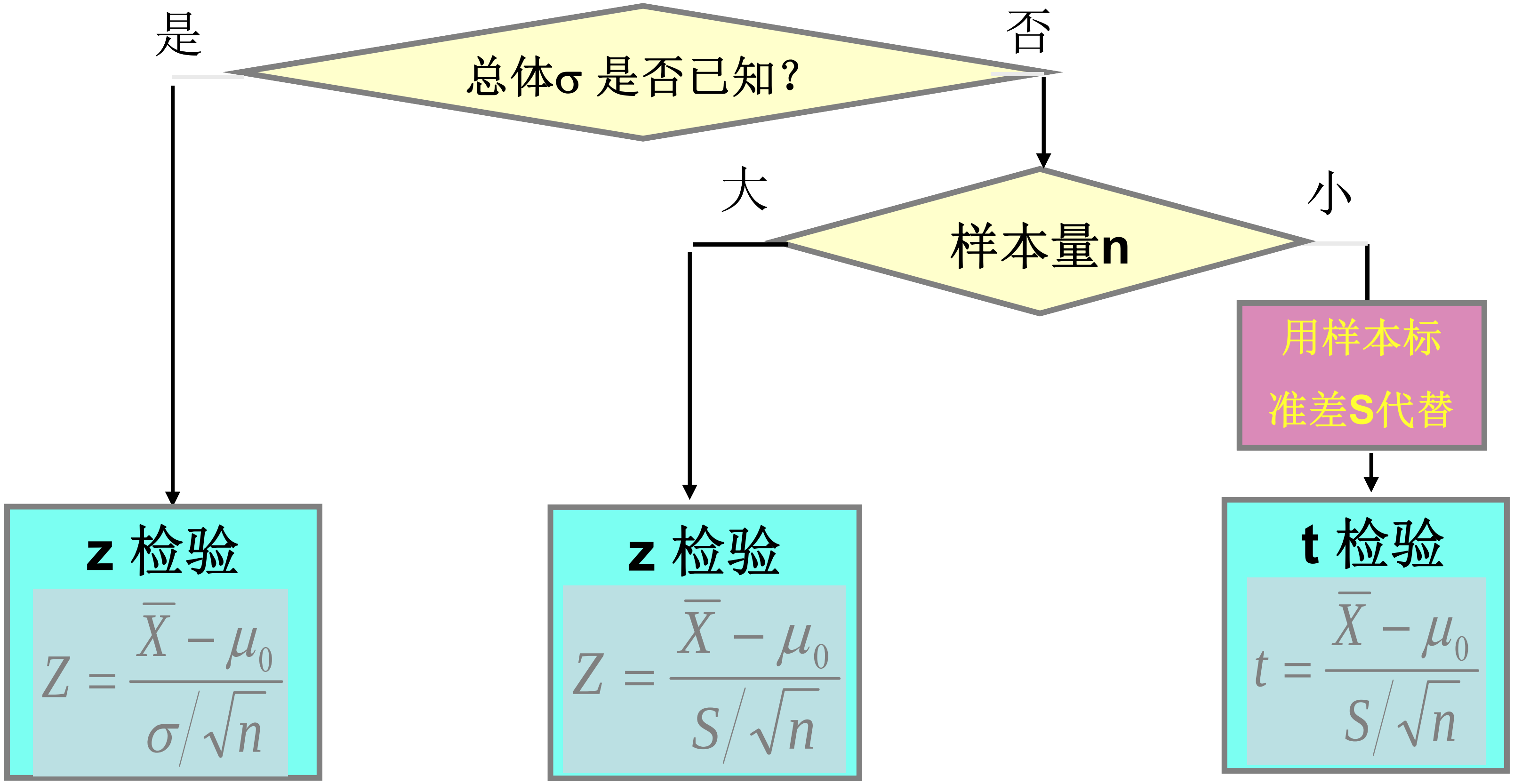
  χ^2检验、

  F-检验

（二）一个总体参数的检验



总体均值的检验
(检验统计量)



总体均值的检验

(σ^2 已知或σ^2未知大样本)

1. 假定条件
   - 总体服从正态分布
   - 若不服从正态分布, 可用正态分布来近似 ($n\geq30$)
2. 使用Z-统计量

□ σ^2 已知：$Z=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$

□ σ^2 未知：$Z=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim N(0,1)$

σ^2 已知均值的检验
(例题分析)

- 【例】某机床厂加工一种零件，根据经验知道，该厂加工零件的椭圆度近似服从正态分布，其总体均值为μ_0=0.081mm，总体标准差为σ= 0.025 。今换一种新机床进行加工，抽取n=200个零件进行检验，得到的椭圆度为0.076mm。试问新机床加工零件的椭圆度的均值与以前有无显著差异？（α＝0.05）

双侧检验

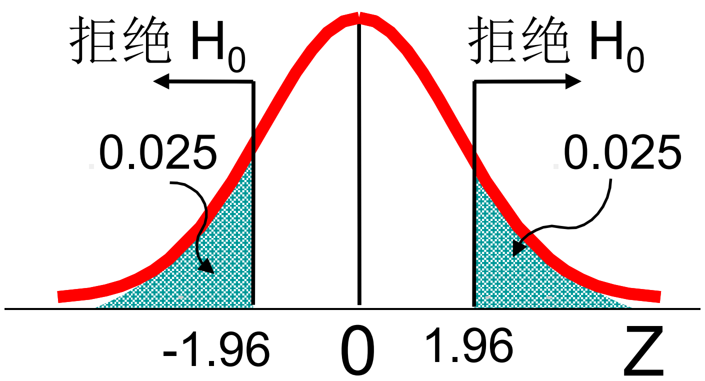
H0: μ = 0.081

H1: μ ≠ 0.081

α = 0.05

***n* = 200**

- **临界值(s):**



检验统计量:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{0.076 - 0.081}{0.025 / \sqrt{200}} = -2.83$$

决策:

在 α = 0.05的水平上拒绝H_0

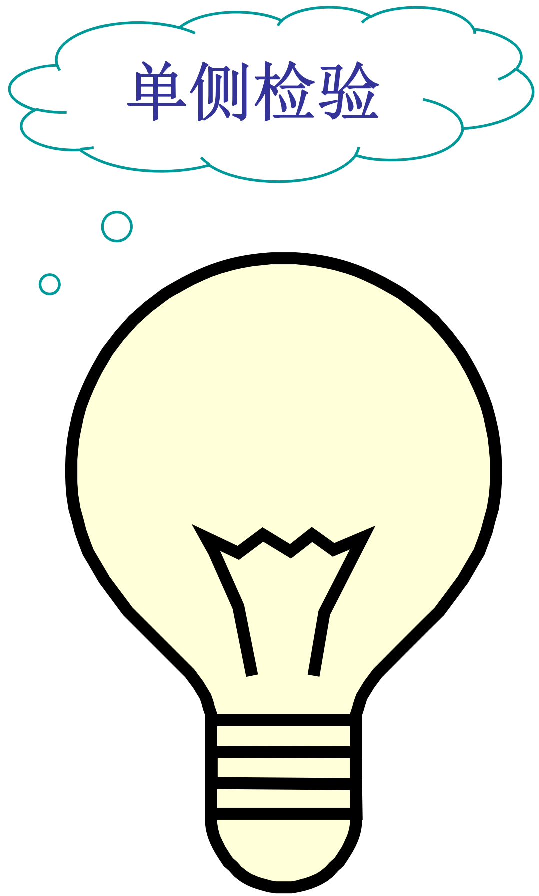
结论:

有证据表明新机床加工的零件的椭圆度与以前有显著差异

σ^2 已知均值的检验

(小样本例题分析)

- 【例】根据过去大量资料，某厂生产的灯泡的使用寿命服从正态分布 N~(1020，100^2)。现从最近生产的一批产品中随机抽取16只，测得样本平均寿命为1080小时。试在0.05的显著性水平下判断这批产品的使用寿命是否有显著提高？(α＝0.05)



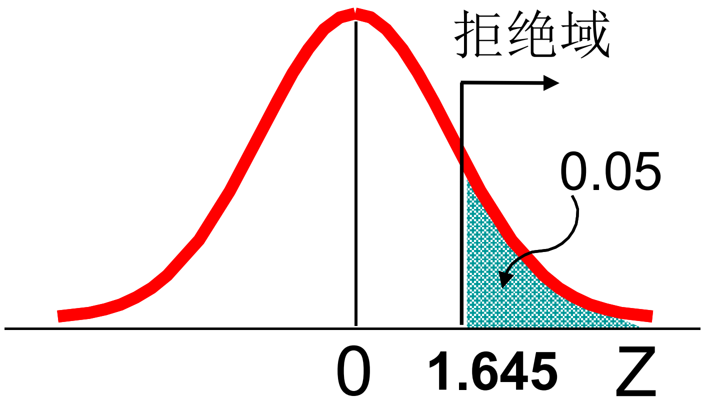
H_0: $\mu \le 1020$

H_1: $\mu > 1020$

$\alpha = 0.05$

$n = 16$

临界值(s):



检验统计量:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1080 - 1020}{100/\sqrt{16}} = 2.4$$

决策:

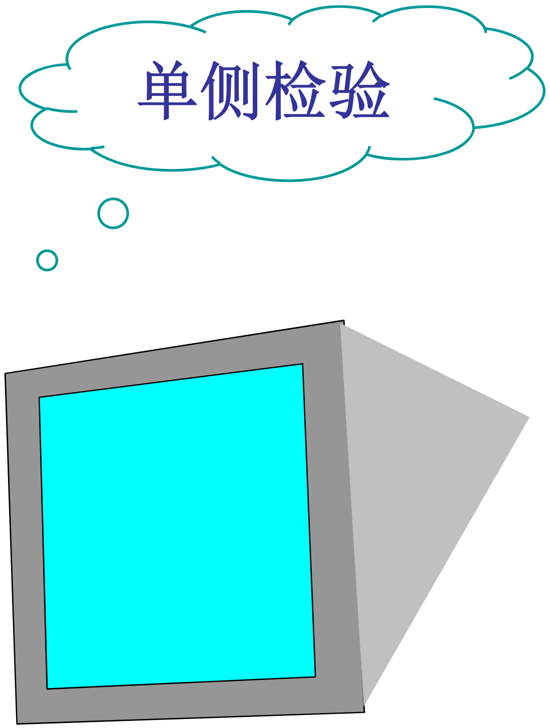
在 $\alpha = 0.05$的水平上拒绝H_0

结论:

有证据表明这批灯泡的使用寿命有显著提高

σ^2 未知大样本均值的检验
(例题分析)

- 【例】某电子元件批量生产的质量标准为平均使用寿命1200小时。某厂宣称他们采用一种新工艺生产的元件质量大大超过规定标准。为了进行验证，随机抽取了100件作为样本，测得平均使用寿命1245小时，标准差300小时。能否说该厂生产的电子元件质量显著地高于规定标准？ ($\alpha=0.05$)



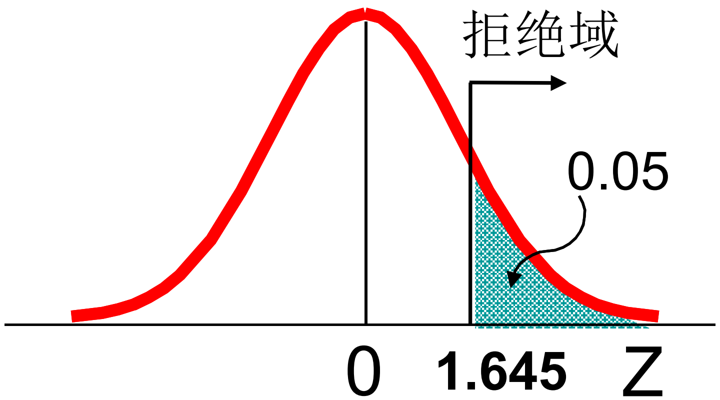
H_0: $\mu \le 1200$

H_1: $\mu > 1200$

$\alpha = 0.05$

$n = 100$

临界值(s):



检验统计量:

$$z = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{1245 - 1200}{300/\sqrt{100}} = 1.5$$

决策:

在 $\alpha = 0.05$的水平上不拒绝H_0

结论:

不能认为该厂生产的元件寿命显著地高于1200小时

总体均值的检验
(σ^2未知、且小样本)

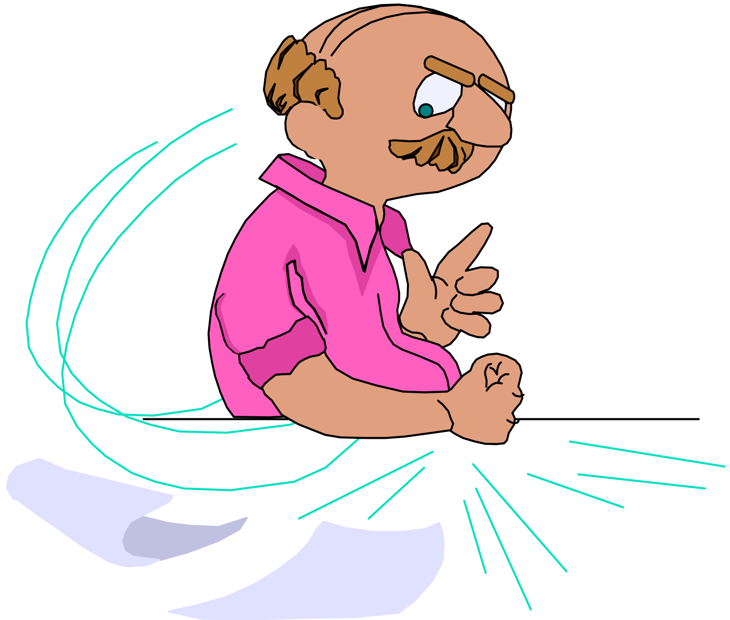
1.假定条件

– 总体为正态分布

– σ ^{2}未知，且小样本

2.使用t 统计量

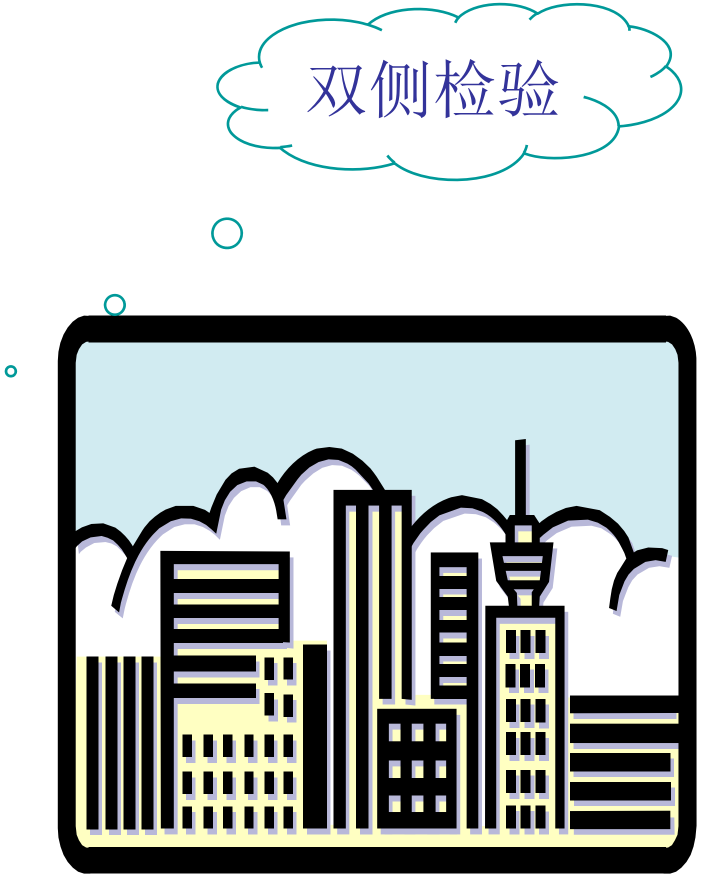
$$t=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$$



σ^2 未知小样本均值的检验

(例题分析)

【例】某机器制造出的肥皂厚度为5cm，今欲了解机器性能是否良好，随机抽取10块肥皂为样本，测得平均厚度为5.3cm，标准差为0.3cm，试以0.05的显著性水平检验机器性能良好的假设。



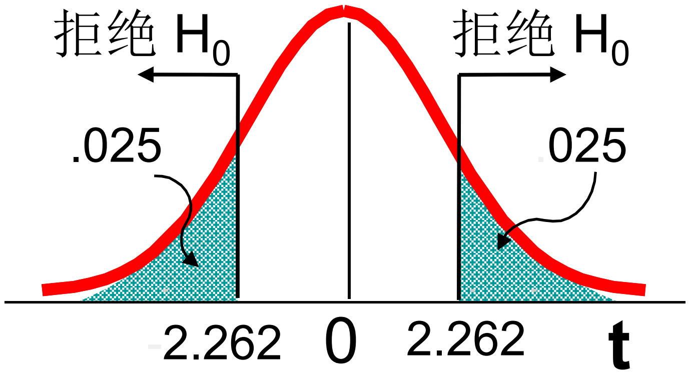
H_0: $\mu = 5$

H_1: $\mu \neq 5$

$\alpha = 0.05$

df = 10 - 1 = 9

临界值(s):



检验统计量:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{5.3 - 5}{0.3/\sqrt{10}} = 3.16$$

决策:

在 $\alpha = 0.05$的水平上拒绝H_0

结论:

说明该机器的性能不好

一个总体比例检验
（大样本）

1. 假定条件
   - 有两类结果
   - 总体服从二项分布
   - 可用正态分布来近似
2. 比例检验的 Z 统计量

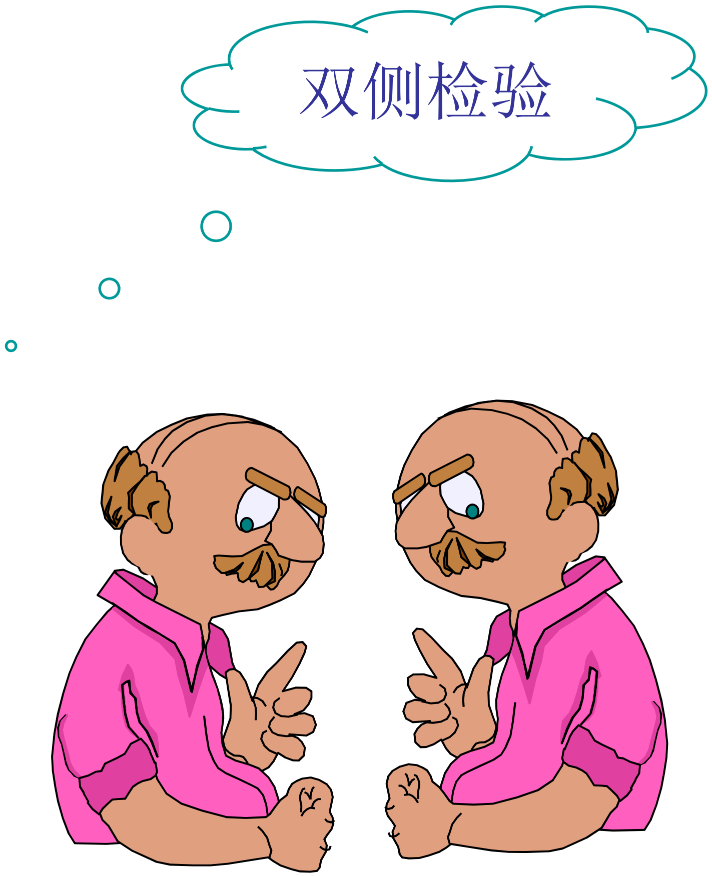
$$Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1)$$

π_0为假设的总体比例

一个总体比例的检验（大样本情形）

(例题分析)

- 【例】一项统计结果声称，某市老年人口（年龄在65岁以上）的比重为14.7%，该市老年人口研究会为了检验该项统计是否可靠，随机抽选了400名居民，发现其中有57人年龄在65岁以上。调查结果是否支持该市老年人口比重为14.7%的看法？(α= **0.05**)



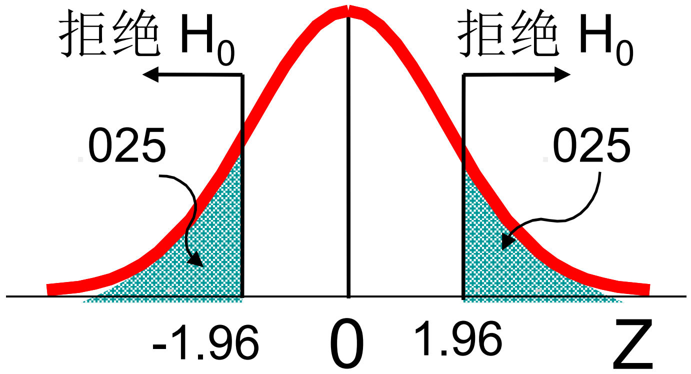
H$_0$: π = 14.7%

H$_1$: $\pi \neq$ 14.7%

α = 0.05

n = 400

临界值(s):



检验统计量:

$$z = \frac{0.1425 - 0.147}{\sqrt{\frac{0.147 \times (1 - 0.147)}{400}}} = -0.254$$

决策:

在α = 0.05的水平上不拒绝H_0

结论:

该市老年人口比重为14.7%

总体方差的检验
(χ^2 检验)

1. 检验一个总体的方差或标准差
2. 假设总体近似服从正态分布
3. 检验统计量

样本方差

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

假设的总体方差

方差的卡方 (χ^2) 检验
(例题分析)

- 【例】某厂商生产出一种新型的饮料装瓶机器，按设计要求，该机器装一瓶一升(1000cm^3)的饮料误差上下不超过1cm^3。如果达到设计要求，表明机器的稳定性非常好。现从该机器装完的产品中随机抽取25瓶，分 别 进 行 测 定 (用 样 本 减 1000cm^3)，得到如下结果。检验该机器的性能是否达到设计要求 (α=0.05)

0.3	-0.4	-0.7	1.4	-0.6
-0.3	-1.5	0.6	-0.9	1.3
-1.3	0.7	1	-0.5	0
-0.6	0.7	-1.5	-0.2	-1.9
-0.5	1	-0.2	-0.6	1.1

双侧检验

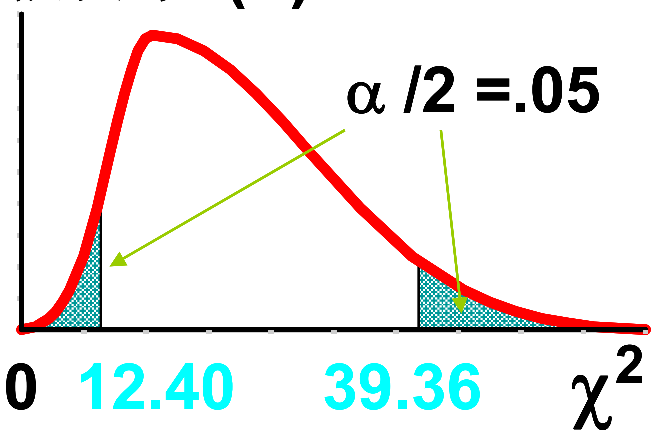
H_0: $\sigma^2 = 1$

H_1: $\sigma^2 \neq 1$

$\alpha = 0.05$

df = 25 - 1 = 24

临界值(s):



统计量:

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

$$= \frac{(25-1)0.866}{1} = 20.8$$

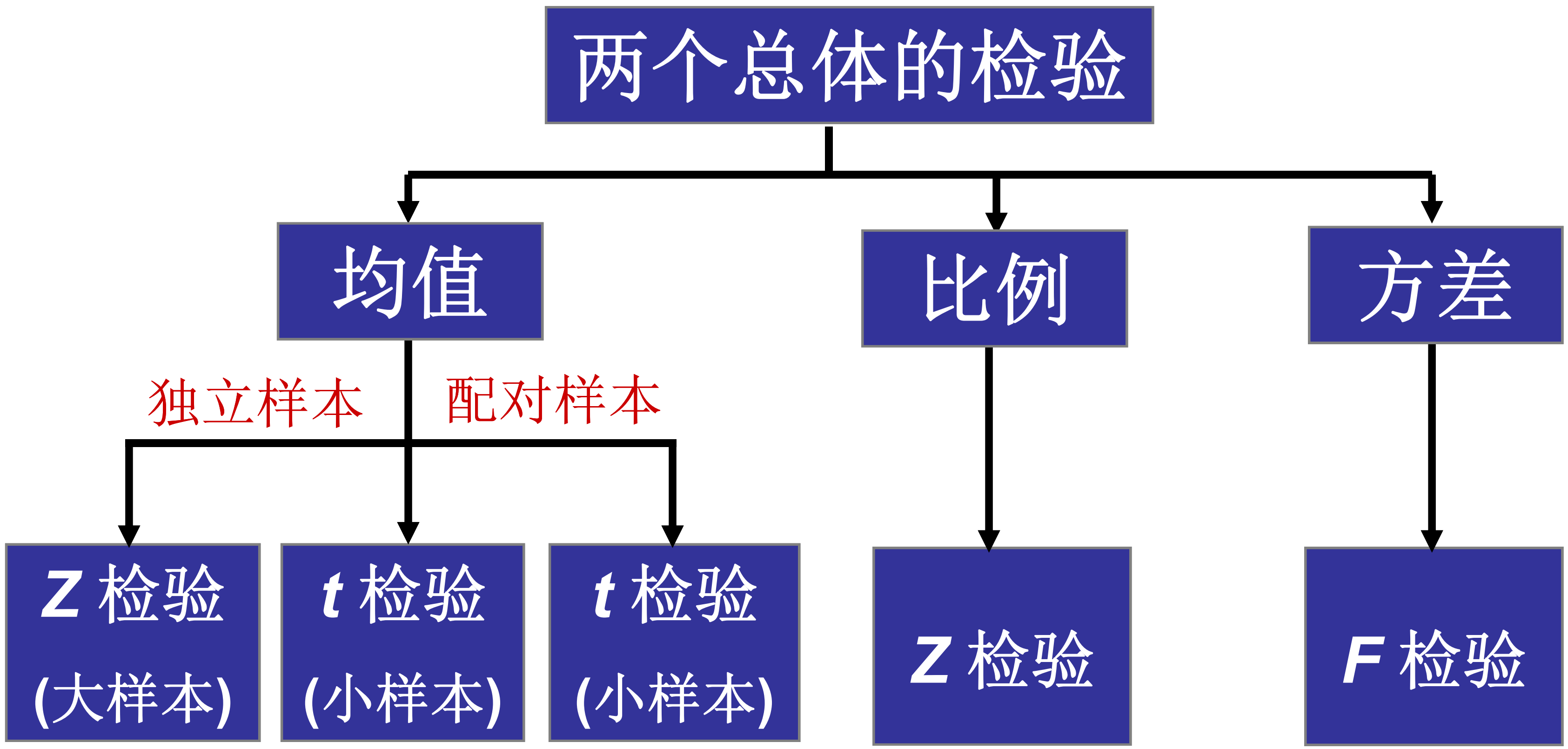
决策:

在 $\alpha = 0.05$的水平上不拒绝H_0

结论:

不能认为该机器的性能未达到设计要求

（三）两个总体参数关系的检验



两个总体均值之差的检验
(σ_1^2、 σ_2^2 已知或未知但大样本)

1. 假定条件
   - 两个样本是独立的随机样本
   - 两个总体都是正态分布
   - 若不是正态分布, 可以用正态分布来近似($n_1 \geq 30$ 和 $n_2 \geq 30$)
2. 检验统计量为

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

两个总体均值之差的检验
(假设的形式)

假设	研究的问题		
	没有差异 有差异	均值$_1$ ≥ 均值$_2$ 均值$_1$ < 均值$_2$	均值$_1$ ≤ 均值$_2$ 均值$_1$ > 均值$_2$
H_0	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \geq 0$	$\mu_1 - \mu_2 \leq 0$
H_1	$\mu_1 - \mu_2 \neq 0$	$\mu_1 - \mu_2 < 0$	$\mu_1 - \mu_2 > 0$

两个总体均值之差的检验

(例题分析)

- 【例】有两种方法可用于制造某种以抗拉强度为重要特征的产品。根据以往的资料得知，第一种方法生产出的产品其抗拉强度的标准差为8公斤，第二种方法的标准差为10公斤。从两种方法生产的产品中各抽取一个随机样本，样本量分别为n_1=32，n_2=40，测得 $\bar{x}_1 = 50$公斤， $\bar{x}_2 = 44$公斤。问这两种方法生产的产品平均抗拉强度是否有显著差别？ ($\alpha = 0.05$)

双侧检验！

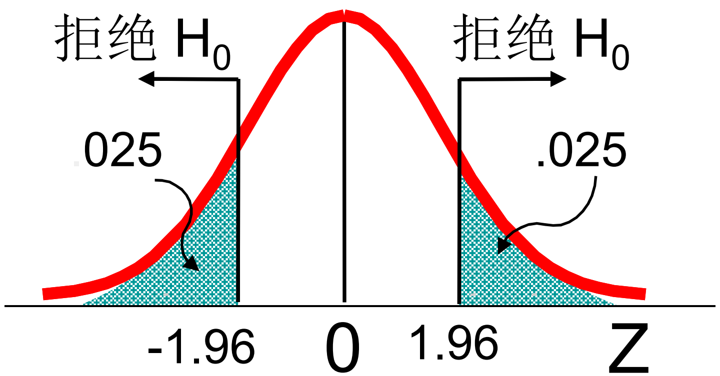
H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

$\alpha = 0.05$

$n_1 = 32$，$n_2 = 40$

临界值(s):



检验统计量:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{50 - 44 - 0}{\sqrt{\frac{64}{32} + \frac{100}{40}}} = 2.83$$

决策:

在 $\alpha = 0.05$的水平上拒绝H_0

结论:

有证据表明两种方法生产的产品其抗拉强度有显著差异

两个总体均值之差的检验
(σ_1^2、 σ_2^2 未知但相等,小样本)

1. 检验具有等方差的两个总体的均值
2. 假定条件
   - 两个样本是独立的随机样本
   - 两个总体都是正态分布
   - 两个总体方差未知但相等$\sigma_1^2 = \sigma_2^2$
3. 检验统计量

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

其中： $S_p^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

两个总体均值之差的检验

(σ_1^2、 σ_2^2 未知且不相等,小样本)

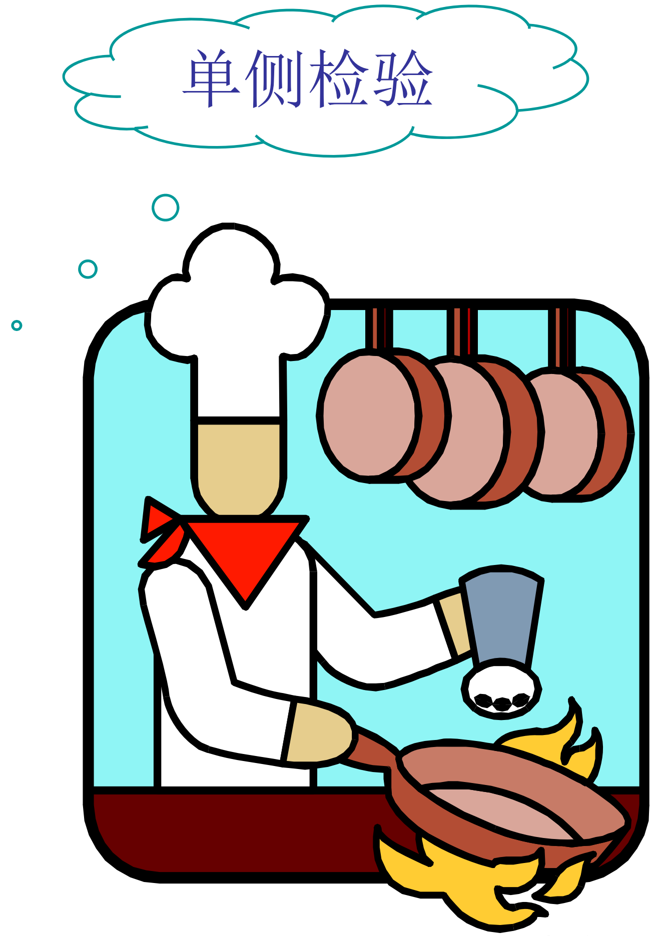
1. 检验具有不等方差的两个总体的均值
2. 假定条件
   - 两个样本是独立的随机样本
   - 两个总体都是正态分布
   - 两个总体方差未知且不相等$\sigma_1^2 \neq \sigma_2^2$
3. 检验统计量

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

两个总体均值之差的检验

(例题分析)

【例】“多吃谷物，将有助于减肥。”为了验证这个假设，随机抽取了35人，询问他们早餐和午餐的通常食谱。根据他们的食谱，将其分为二类，一类为经常的谷类食用者(总体1，15人)，一类为非经常谷类食用者(总体2，20人)。然后测度每人午餐的大卡摄取量。经过一段时间的实验，得到如下结果：检验该假设 ($\alpha = 0.05$)



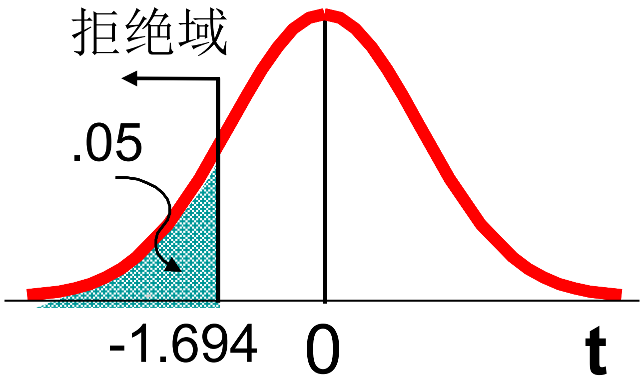
H_0: $\mu_1 - \mu_2 \geq 0$

H_1: $\mu_1 - \mu_2 < 0$

$\alpha = 0.05$

$n_1 = 15$，$n_2 = 20$

临界值(s):



检验统计量:

$$t = \frac{583 - 629.25}{\sqrt{\frac{2431.429}{15} + \frac{3675.461}{20}}} = -2.4869$$

决策:

在 $\alpha = 0.05$的水平上拒绝H_0

结论:

没有证据表明多吃谷物将有助于减肥

两个总体均值之差的检验
(匹配样本的 *t* 检验)

- 1. 检验两个总体的均值
  - 配对或匹配
  - 重复测量 (前/后)
- 3. 假定条件
  - 两个总体都服从正态分布
  - 如果不服从正态分布，可用正态分布来近似 ($n_1 \geq 30$, $n_2 \geq 30$)

匹配样本的 *t* 检验
(假设的形式)

假设	研究的问题		
	没有差异 有差异	总体$_1$ ≥ 总体$_2$ 总体$_1$ < 总体$_2$	总体$_1$ ≥ 总体$_2$ 总体$_1$ > 总体$_2$
H_0	$\mu_D = 0$	$\mu_D \geq 0$	$\mu_D \leq 0$
H_1	$\mu_D \neq 0$	$\mu_D < 0$	$\mu_D > 0$

注：$D_i = X_{1i} - X_{2i}$，对第 i 对观察值

匹配样本的 *t* 检验

(数据形式)

观察序号	样本1	样本2	差值
1	x_{11}	x_{21}	$D_1 = x_{11} - x_{21}$
2	x_{12}	x_{22}	$D_1 = x_{12} - x_{22}$
⋮	⋮	⋮	⋮
i	x_{1i}	x_{2i}	$D_1 = x_{1i} - x_{2i}$
⋮	⋮	⋮	⋮
n	x_{1n}	x_{2n}	$D_1 = x_{1n} - x_{2n}$

匹配样本的 *t* 检验
(检验统计量)

统计量

$$t = \frac{\bar{X}_D - D_0}{S_D / \sqrt{n_D}}$$

D_0：假设的差值

自由度$df = n_D - 1$

样本差值均值

$$\bar{X}_D = \frac{\sum_{i=1}^{n} D_i}{n_D}$$

样本差值标准差

$$S_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \bar{X}_D)^2}{n_D - 1}}$$

匹配样本的 t 检验

(例题分析)

- 【例】一个以减肥为主要目标的健美俱乐部声称，参加其训练班至少可以使减肥者平均体重减重**8.5**kg以上。为了验证该宣称是否可信，调查人员随机抽取了**10**名参加者，得到他们的体重记录如下表：

训练前	94.5	101	110	103.5	97	88.5	96.5	101	104	116.5
训练后	85	89.5	101.5	96	86	80.5	87	93.5	93	102

在 $\alpha = 0.05$ 的显著性水平下，调查结果是否支持该俱乐部的声称？

单侧检验

样本差值计算表		
训练前	训练后	差值D_i
94.5	85	9.5
101	89.5	11.5
110	101.5	8.5
103.5	96	7.5
97	86	11
88.5	80.5	8
96.5	87	9.5
101	93.5	7.5
104	93	11
116.5	102	14.5
合计	—	98.5

差值均值 $$\bar{x}_D = \frac{\sum_{i=1}^{n} D_i}{n_D} = \frac{98.5}{10} = 9.85$$

差值标准差 $$s_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \bar{x}_D)^2}{n_D - 1}} = \sqrt{\frac{43.525}{10-1}} = 2.199$$

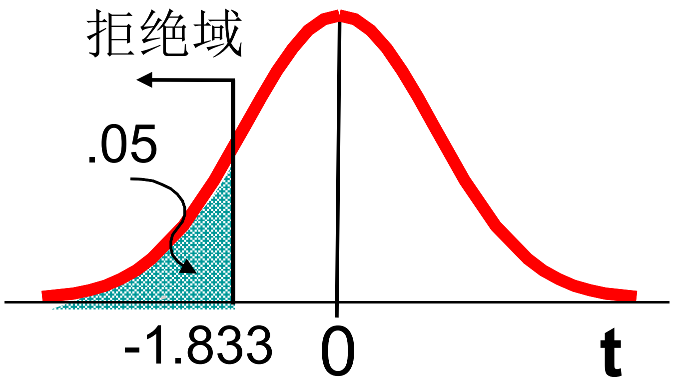
H_0: $\mu_1 - \mu_2 \geq 8.5$

H_1: $\mu_1 - \mu_2 < 8.5$

$\alpha = 0.05$

df = 10 - 1 = 9

临界值(s):



检验统计量:

$$t = \frac{\bar{x}_D - D_0}{s_D / \sqrt{n_D}} = \frac{9.85 - 8.5}{2.199 / \sqrt{10}} = 1.9413$$

决策:

在 $\alpha = 0.05$的水平上不拒绝H_0

结论:

不能认为该俱乐部的宣称不可信

两个总体比例之差的Z检验

1. 假定条件
   - 两个总体是独立的
   - 两个总体都服从二项分布
   - 可以用正态分布来近似
2. 检验统计量

$$Z = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{\dfrac{P_1(1-P_1)}{n_1} + \dfrac{P_2(1-P_2)}{n_2}}} \sim N(0,1)$$

两个总体比例之差的检验
(假设的形式)

假设	研究的问题		
	没有差异 有差异	比例$_1$ ≥比例$_2$ 比例$_1$ < 比例$_2$	总体$_1$ ≤比例$_2$ 总体$_1$ > 比例$_2$
H_0	$\pi_1-\pi_2 = 0$	$\pi_1-\pi_2 \geq 0$	$\pi_1-\pi_2 \leq 0$
H_1	$\pi_1-\pi_2 \neq 0$	$\pi_1-\pi_2 < 0$	$\pi_1-\pi_2 > 0$

两个总体比例之差的Z检验
(例题分析)

【例】对两个大型企业青年工人参加技术培训的情况进行调查，调查结果如下：甲厂：调查**60**人，**18**人参加技术培训。乙厂调查**40**人，**14**人参加技术培训。能否根据以上调查结果认为乙厂工人参加技术培训的人数比例高于甲厂？(**α = 0.05**)

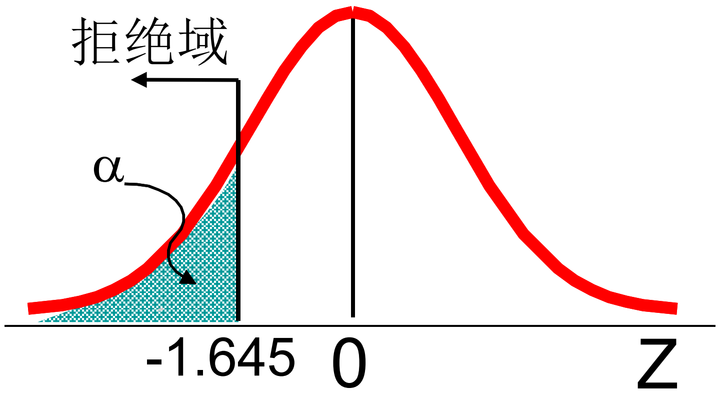
H_0: $\pi_1 - \pi_2 \geq 0$

H_1: $\pi_1 - \pi_2 < 0$

$\alpha = 0.05$

$n_1 = 60$，$n_2 = 40$

临界值(s):



检验统计量:

$$z = \frac{0.30 - 0.35 - 0}{\sqrt{\frac{0.30(1-0.30)}{60} + \frac{0.35(1-0.35)}{40}}} = -0.52$$

决策:

在 $\alpha = 0.05$的水平上不拒绝H_0

结论:

没有证据表明乙厂工人参加技术培训的人数比例高于甲厂

两个总体方差比的检验
(F 检验)

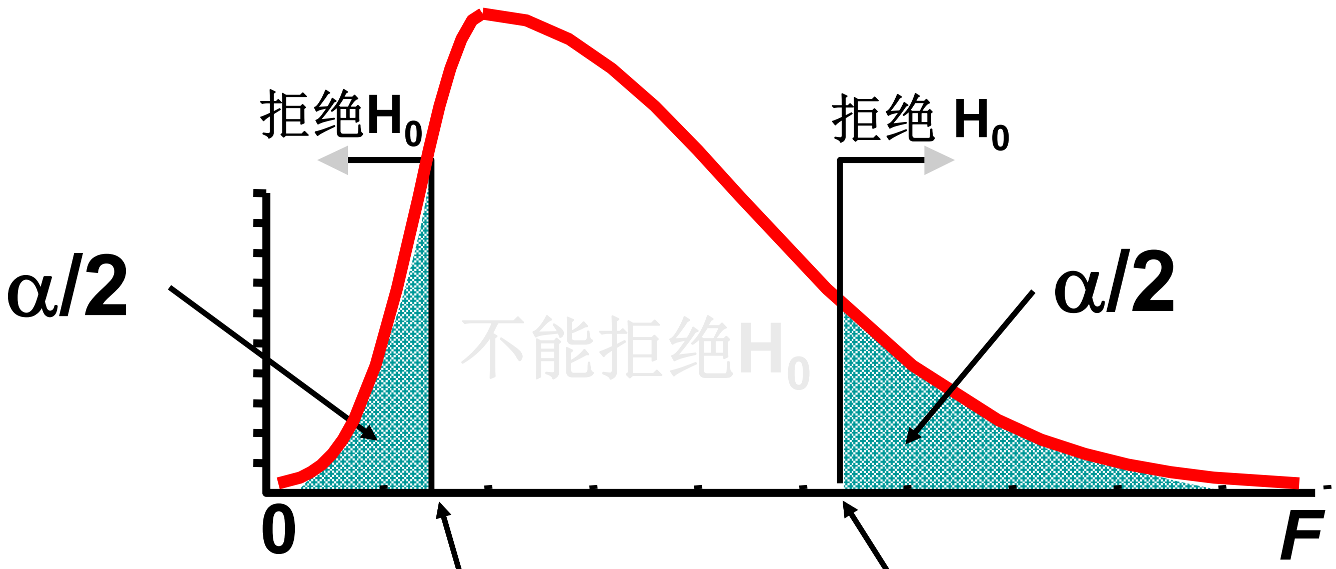
1. 假定条件
   - 两个总体都服从正态分布
   - 两个独立的随机样本
2. 假定形式

   H_0：$\sigma_1^2 = \sigma_2^2$　或　H_0：$\sigma_1^2 \geq \sigma_2^2$ (或 $\leq$)

   H_1：$\sigma_1^2 \neq \sigma_2^2$　　　H_1：$\sigma_1^2 < \sigma_2^2$ (或 $>$)
3. 检验统计量

   $F =$ （S_1^2/σ_1^2 ）$/S_2^2/\sigma_2^2 \sim F(n_1 - 1，n_2 - 1)$

两个总体方差的 *F* 检验
(临界值)



$$F_{1-\alpha/2}(n_1-1,n_2-1)=\frac{1}{F_{\alpha/2}(n_2-1,n_1-1)}$$

$$F_{\alpha/2}(n_1-1,n_2-1)$$

两个总体方差的 *F* 检验
(例题分析)

- 【例】生产过程的方差可以衡量生产质量的优劣。现有一家销售公司欲从两家厂商选择进货来源，以两家工厂生产的同类产品的重量方差为选择依据。重量方差越小，说明质量越稳定，该公司将选取该厂商为进货来源。分别抽取两家厂商8个批次的产品，测得样本数据如下（假设产品重量都服从正态分布）：

两厂商同类产品不同批次的平均重量（单位：克）

批次	1	2	3	4	5	6	7	8
甲厂商	100	102	98	96.8	103.5	101	97.4	100
乙厂商	109.2	101	100	102	98.5	96.2	101.5	94.1

在5%显著性水平下，试问甲厂商生产的产品方差是否显著小于乙厂商？可否选择甲作为进货来源？

解：依题意，本题检验的对象为甲厂商产品方差是否显著小于乙厂商，因此采用单侧检验。

如前所述，我们令乙厂商抽取的产品样本来源于总体1，甲厂商的产品样本来源于总体2；且 $n_1 = 8, n_2 = 8$ 通过计算得到两个样本的标准差：

$S_1 = 4.51, S_2 = 2.33$

（1）建立原假设和备选假设：

$$H_0:\sigma_1^2 \le \sigma_2^2, H_1:\sigma_1^2 > \sigma_2^2$$

（2）查表得临界值并确定拒绝域：

从 F分布表中查得临界值为 $F_\alpha(n_1-1, n_2-1) = 3.79,$ 拒绝域为[3.79，∞）

（3）样本统计量的计算及判断：

在H_0为真的条件下，由样本数据计算*F*统计量的值，为

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{4.51^2}{2.33^2} = 3.75 \le 3.79$$

由于F_0 未落入拒绝域，因此不能拒绝原假设 $H_0 : \sigma_1^2 \le \sigma_2^2$ ，不能肯定甲厂商生产的产品平均重量方差较小，质量更为稳定。

假设检验和置信区间

- 假设检验与置信区间有什么区别和联系呢?
  – 从数学上来说，假设检验与置信区间是对偶的。一般地，如果置信区间不包含假设的特定值，则拒绝 。如果置信区间包含假设的取值，则不能拒绝 。所以，对假设检验而言，“非拒绝域”等同于假设的总体取值落入置信区间。

– 在许多方面，置信区间比假设检验提供的信息要多。置信区间给了我们一个参数值的可能范围，而假设检验只考虑到一个可能值。
– 例如在假设检验中，如果总体参数不是100，我们就不清楚它是多少了。有时这一个值是非常重要或有意义的，像检验两均值之差是否等于零。可是即使当我们拒绝了零假设，并得到均值之差不为零的结论之后，紧接着的一个问题就是该差异是多大。这个问题可以由置信区间来回答的。

- “尽管人们可能更希望得到置信区间，假设检验还是被广泛地应用于大多数领域中。主要原因就是统计软件很多时候并不自动计算置信区间。”

这一观点你是否认同？