

第六章

参数估计与假设检验



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE

目录



第一节 参数估计



第二节 假设检验



第三节 假设检验中的两类错误



第四节 假设检验与区间估计的关系





重点难点

掌握

抽样误差的概念、**均数的标准误**和率的标准误、参数估计的方法（**区间估计**）；假设检验的基本原理和基本步骤，以及检验的两类错误。

熟悉

标准误及参数估计的计算公式、假设检验两类错误的关系、假设检验与区间估计的关系。



第一节

参数估计



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE



参数估计

指由**样本统计量**估计**总体参数**，是统计推断的重要内容之一。

包括：

- 点估计：**样本均值估计总体均值，样本标准差估计总体标准差。**
- 区间估计：**按预先给定的概率计算出一个区间，能够包含未知的总体参数。**





一、抽样误差

- **抽样误差的概念**：在医学研究中，绝大多数情况是抽样后由样本数据推断总体特征。由于个体存在差异，因此通过随机抽样得到的样本在推论总体时会存在一定的误差（如样本均数 $\bar{x}$ 往往不等于总体均数 μ ），这种由抽样造成的样本统计量与总体参数之间的差异称为抽样误差（sampling error）。

举例：假设健康成年男子的红细胞数服从均数为 $\mu=4.75 \times 10^{12} / L$ ，标准差为 $\sigma=0.38 \times 10^{12} / L$ 的正态分布，现随机抽取140人，计算红细胞的样本均数为 $\bar{X} = 4.77 \times 10^{12} / L$ ，造成样本均数与总体均数不相等的原因即为抽样误差。





可以想到，再随机抽取140名成年男子进行测量，其平均红细胞数会是另一个结果，对于抽样研究，抽样误差不可避免。

问题：

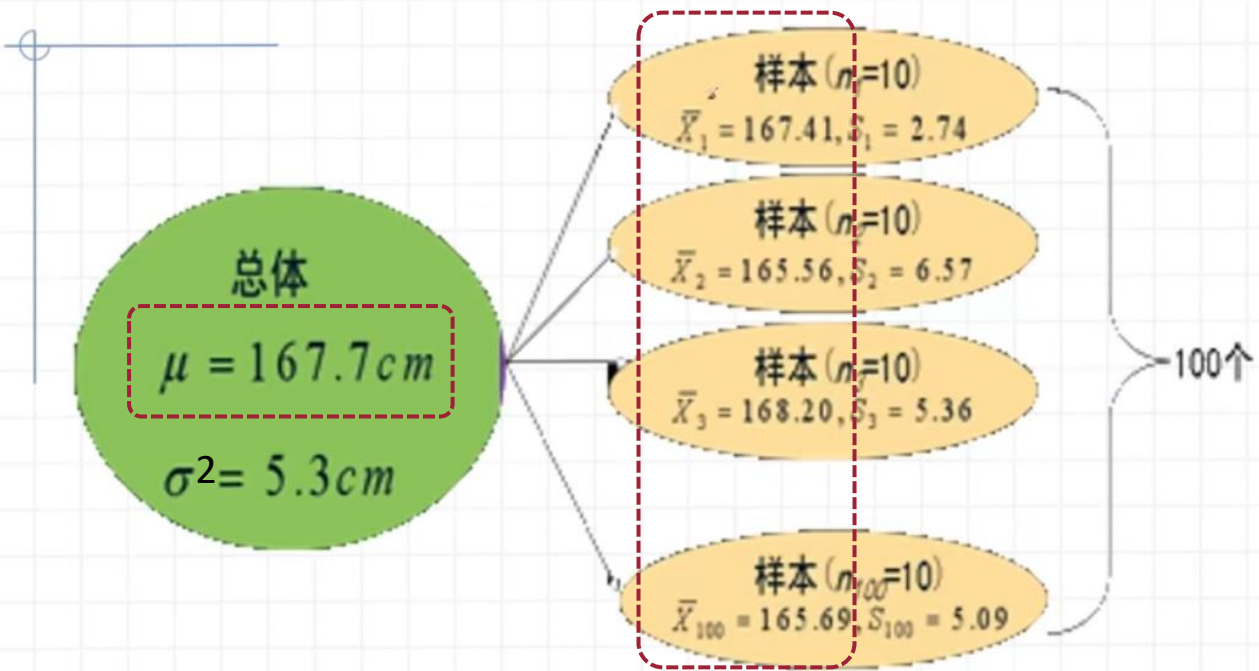
- 如何能精确地估计成年男子红细胞数的总体均数？
- 如何评价估计值的准确程度？

了解抽样误差的分布规律，可以帮助我们知道估计值的准确程度。





例1 若某市2021年18岁男生身高服从均数 μ 、标准差 σ 的正态分布。从该正态分布 $N(167.7, 5.32)$ 总体中随机抽样，共抽了100次，每次样本含量 $n_j=10$ 人，得到每个样本均数及标准差如下图



原因：均数的抽样误差

如何判断判断均数抽样误差的大小？

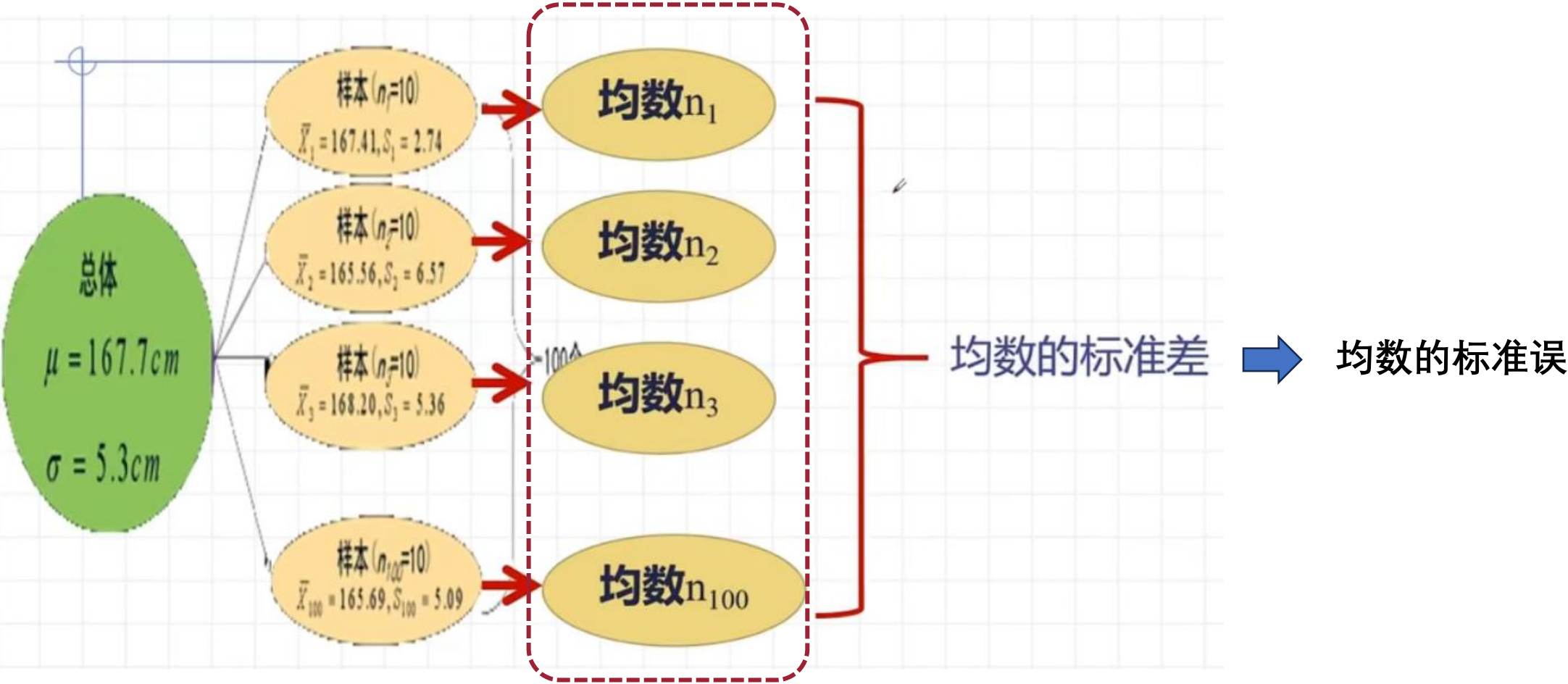


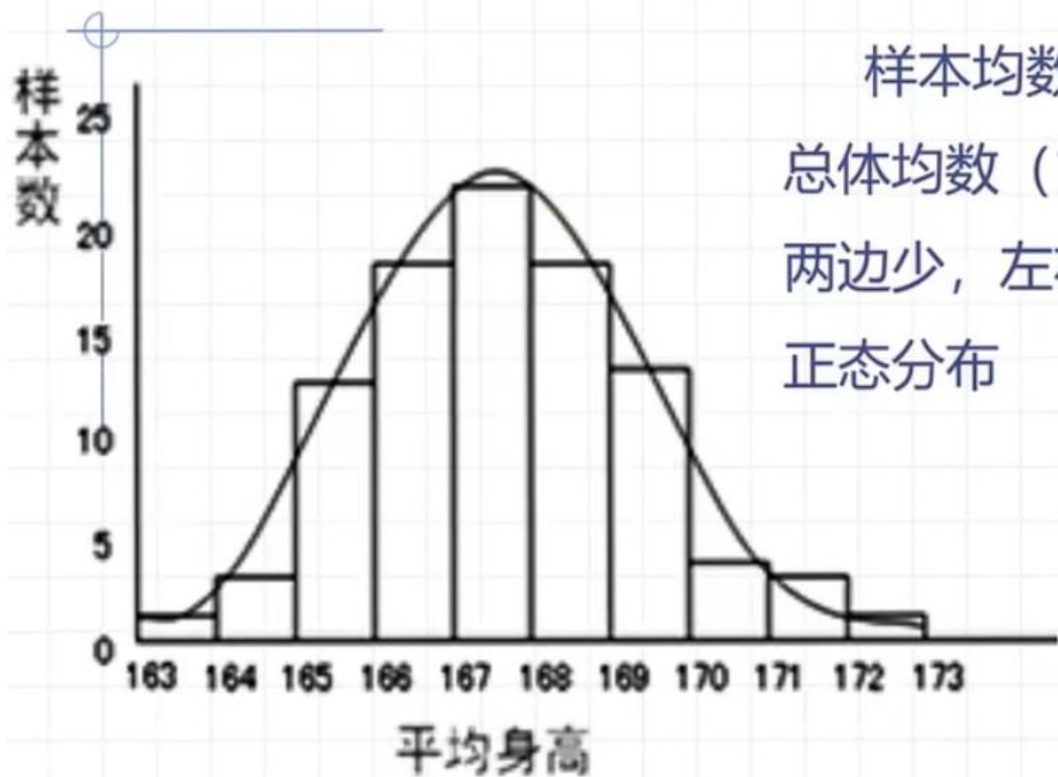
标准误

1. 定义

- 统计学中为了区别个体观察值之间变异的标准差与反映样本均数之间变异的标准差，将后者称为均数的标准误（standard error of mean）。显然，标准误小于原始测量值的标准差，标准误越小说明估计越精确，因此可以用均数的标准误表示均数抽样误差的大小。







样本均数的分布特征：围绕着总体均数（167.7cm），中间多，两边少，左右基本对称，也服从正态分布





若从正态分布总体 $N(\mu, \sigma^2)$ 中，反复多次随机抽取样本含量固定为n的样本，那么这些**样本均数** $\bar{X}$ 也服从正态分布，即 $\bar{X}$ 的总体均数仍为 μ ， $\bar{X}$ 的标准差为 $\sigma/\sqrt{n}$ 。

事实上，在样本含量n很大的情况下（如 $n \geq 50$ ），无论原始测量变量服从什么分布， $\bar{X}$ 的抽样分布都近似服从正态分布 $N(\mu, \sigma^2/n)$ ，这就是**中心极限定理**。

抽样次数	样本1取值	样本2取值	样本3取值	均值
1	1	1	1	1
2	1.1	1.2	1.3	1.2
...	...	...	...	...
n	1.5	1.4	1.6	1.4

$N(\mu, \sigma^2/n)$

☑ 中心极限定理:

- 当样本含量很大的情况下, 无论原始测量变量服从什么分布, $\bar{X}$ 的抽样分布均近似服从正态分布。

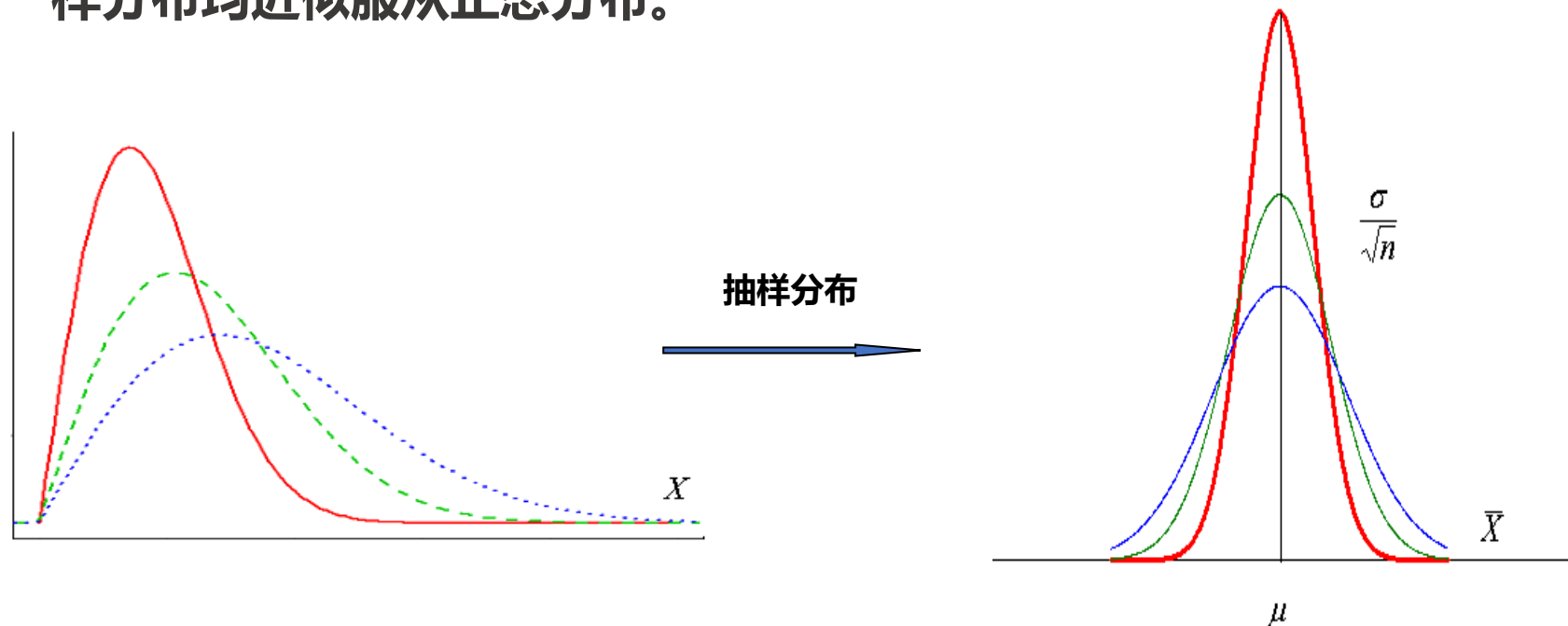
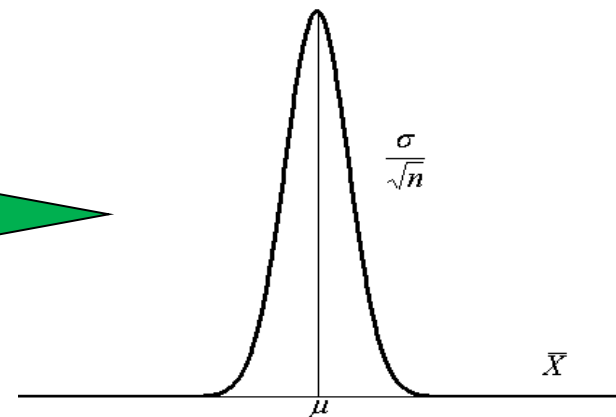
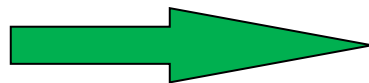
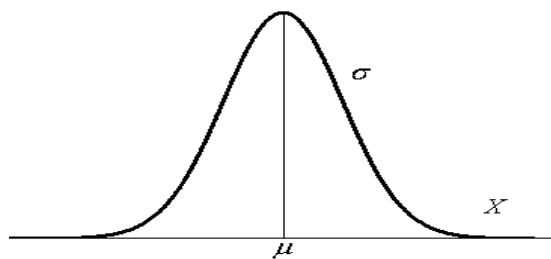


图1 抽样分布与中心极限定理示意图

2. 标准误的计算公式

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$



其中，符号 $\sigma_{\bar{X}}$ 表示均数的标准误，由上式可见，在**样本含量一定**的情况下，均数的标准误与标准差成正比，说明当总体中各观测值变异较小时，抽到的样本均数 $\bar{X}$ 与总体均数 μ 可能相差较小，用 $\bar{X}$ 估计 μ 的可靠程度较高；当总体中各观测值变异较大时，抽到的样本均数 $\bar{X}$ 与总体均数 μ 可能相差很大，用 $\bar{X}$ 估计 μ 的可靠程度也相对较低。

均数的标准误与样本含量的平方根 $\sqrt{n}$ 成反比，说明在同一总体中随机抽样，样本含量 n 越大，标准误越小。



3. 均数的标准误的估计

- 实际中，总体标准差 σ 往往未知，通常用样本标准差 S 估计 σ ，求得样本均数标准误的估计值 $S_{\bar{x}}$ ，计算公式为：

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

例6-1 在某地随机抽查成年男子140人，得红细胞均数为 $4.77 \times 10^{12}/L$ ，标准差为 $0.38 \times 10^{12}/L$ ，试计算其标准误。按公式计算得：

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{0.38}{\sqrt{140}} = 0.032 (\times 10^{12} / L)$$



3. 均数的标准误的用途

1. 可用来衡量**样本均数的可靠性**

标准误



抽样误差



均数间的差异



样本均数估计总体均数

2. 与样本均数结合, 可用于估计**总体均数的可信 (置信) 区间**

3. 可用于进行**均数的假设检验**



说明

标准差与标准误的比较

标 准 差		标 准 误
含 义	描述正态分布资料的离散趋势, 反映个体变量值的变异程度	描述均数的抽样误差, 反映样本均数对总体均数的代表性
计算公式	总体标准差 σ	理论值 $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
	样本标准差 $s = \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n - 1}}$	估计值 $s_{\bar{X}} = \frac{s}{\sqrt{n}}$
与样本含量的关系	样本量越大, 样本标准差越稳定, 越接近总体标准差, 即 $s \rightarrow \sigma$	样本量越大, 标准误越小, 当 $n \rightarrow \infty$ 时, $\sigma_{\bar{X}}$ (或 $s_{\bar{X}}$) $\rightarrow 0$ 。
用 途	用于描述性分析: 描述离散趋势; 计算变异系数; 估计频数分布; 确定医学参考值范围; 用于计算标准误	用于反映抽样误差大小 用于统计推断: 估计总体均数的可信区间 假设检验





(二) 率的标准误

1. 定义

样本^率的标准差称为率的标准误 (standard error of rate) , 可用来描述样本率抽样误差的大小。率的标准误越小, 则率的抽样误差越小, 率的标准误越大, 则率的抽样误差越大。

现记^{总体率参数}为 π , ^{样本率}为 p ; 在 n 次独立重复试验中出现的“阳性”次数记为 X , 则样本率为: $p=X/n$.

样本率 p 的总体方差为 $\sigma_p^2 = \frac{\pi(1-\pi)}{n}$

p 的标准差 (即率的标准误) 为 $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$





2. 率的标准误的估计

在一般情况下，总体率 π 往往是未知的，此时可用样本率 p 来估计总体率 π ，则 σ_p 的估计值为：

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \qquad S_p = \sqrt{\frac{p(1-p)}{n}}$$





总结

中心极限定理

事实上, 在样本含量 n 很大的情况下 (如 $n \geq 50$), 无论原始测量变量服从什么分布, $\bar{X}$ 的抽样分布都近似服从正态分布 $N(\mu, \sigma^2/n)$, 这就是中心极限定理。

均数的标准误

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

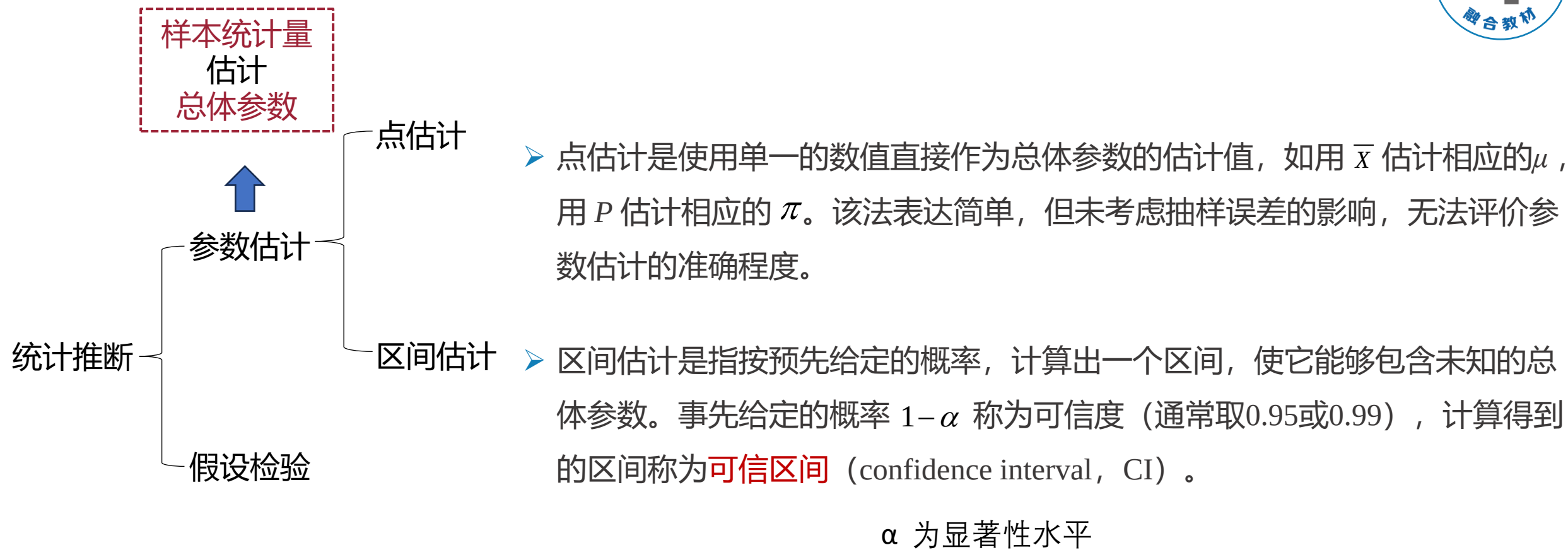
$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

率的标准误

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

$$S_p = \sqrt{\frac{p(1-p)}{n}}$$







可信区间的两个要素

- 准确度：由可信度 $1-\alpha$ 来反映，可信度越大，准确度越高。
- 精密度：由区间的宽度来反映，宽度越小，精密度越大。

问题一：希望知道某高校男生的平均身高，给出它的可信区间的范围为（负无穷，正无穷），请问它的准确度是多少？

问题二：知道某高校男生的平均身高为175，根据三个样本得出给出三个的可信区间，分为为（170，180）、（173，176）、（160，180），请问哪个精密度更高？

在样本含量一定的情况下，二者是矛盾的。若提高可信度，则估计的区间变宽，精确度降低。

可信度确定的情况下，增加样本量可缩小区间宽度。



三、总体均数的区间估计

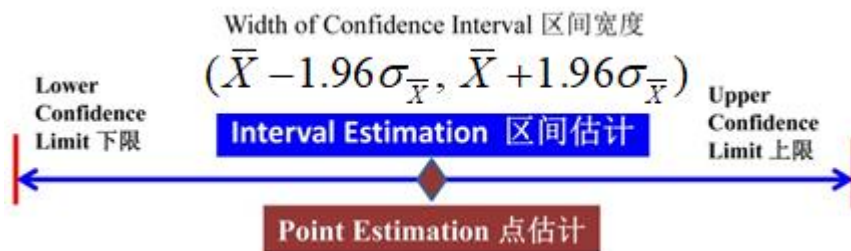
(一) σ 已知

- 如果变量 X 服从均数为 μ 、标准差为 σ 的正态分布, $\bar{X}$ 的抽样分布都近似服从正态分布 $N(\mu, \sigma^2/n)$, 则 $z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ 服从**标准正态分布**。

$$P\left(-1.96 \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq 1.96\right) = 0.95 \quad \longrightarrow \quad P\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

则总体均数95%可信区间为:

$$(\bar{X} - 1.96\sigma_{\bar{X}}, \bar{X} + 1.96\sigma_{\bar{X}})$$



**解读**

说明区间是固定的，参数是变化的，所以可能落在区间或者不落入区间。（实际 μ 是常数，不变的）

总体参数 μ 是客观存在的常数，对置信区间描述正确的是：

- 总体参数落在置信区间的概率为 $1-\alpha$ 。 ✗
- 每一置信区间以95%的置信度包含总体参数 μ 。 ✓

说明总体参数不变，区间在变，区间可能包含总体参数，可能不包含总体参数

总体均数估计的95%可信区间表示该区间包含总体均数 μ 的概率为95%，即若作100次抽样，算得100个可信区间，则平均有95个可信区间包括 μ （估计正确）。





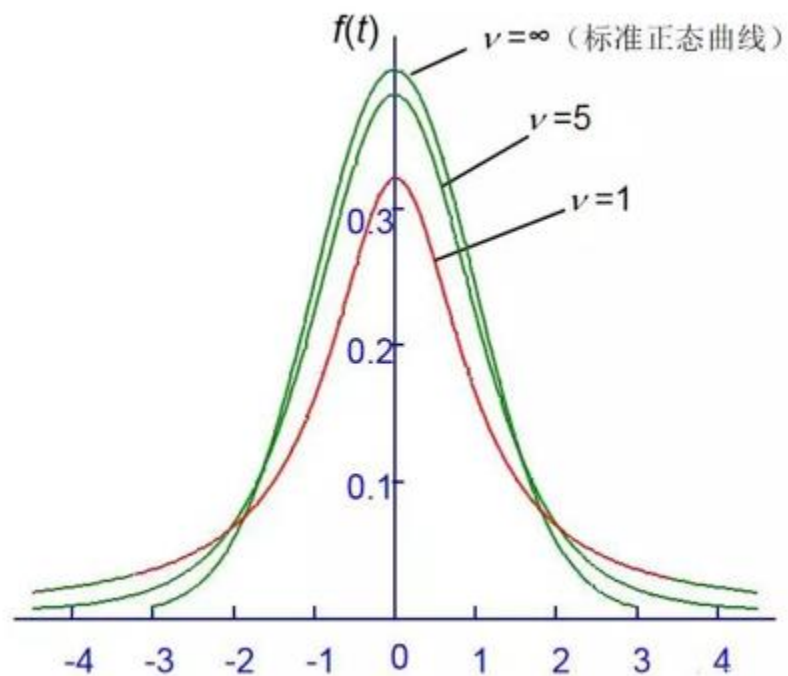
(二) σ 未知

1. t 分布

- 事实上, 总体标准差 σ 通常是未知的, 这时我们可以用其估计量 S 代替 σ
- 但在这种情况下, $(\bar{X} - \mu) / (S / \sqrt{n})$ 已不再服从标准正态分布, 而是服从自由度为 $\nu = n - 1$ 的 t 分布。 t 分布是以 0 为中心的对称分布。

$$t = \frac{\bar{X} - \mu}{S_{\bar{X}}} = \frac{\bar{X} - \mu}{S / \sqrt{n}} \quad \text{自由度 } \nu = n - 1$$





不同自由度的 t 分布图



William Gosset

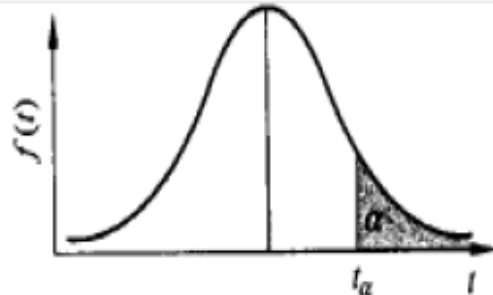
t分布曲线下面积及95%的界值不是一个常量，随自由度大小而变化，随着自由度的增大，t分布曲线越来越接近于标准正态分布曲线。

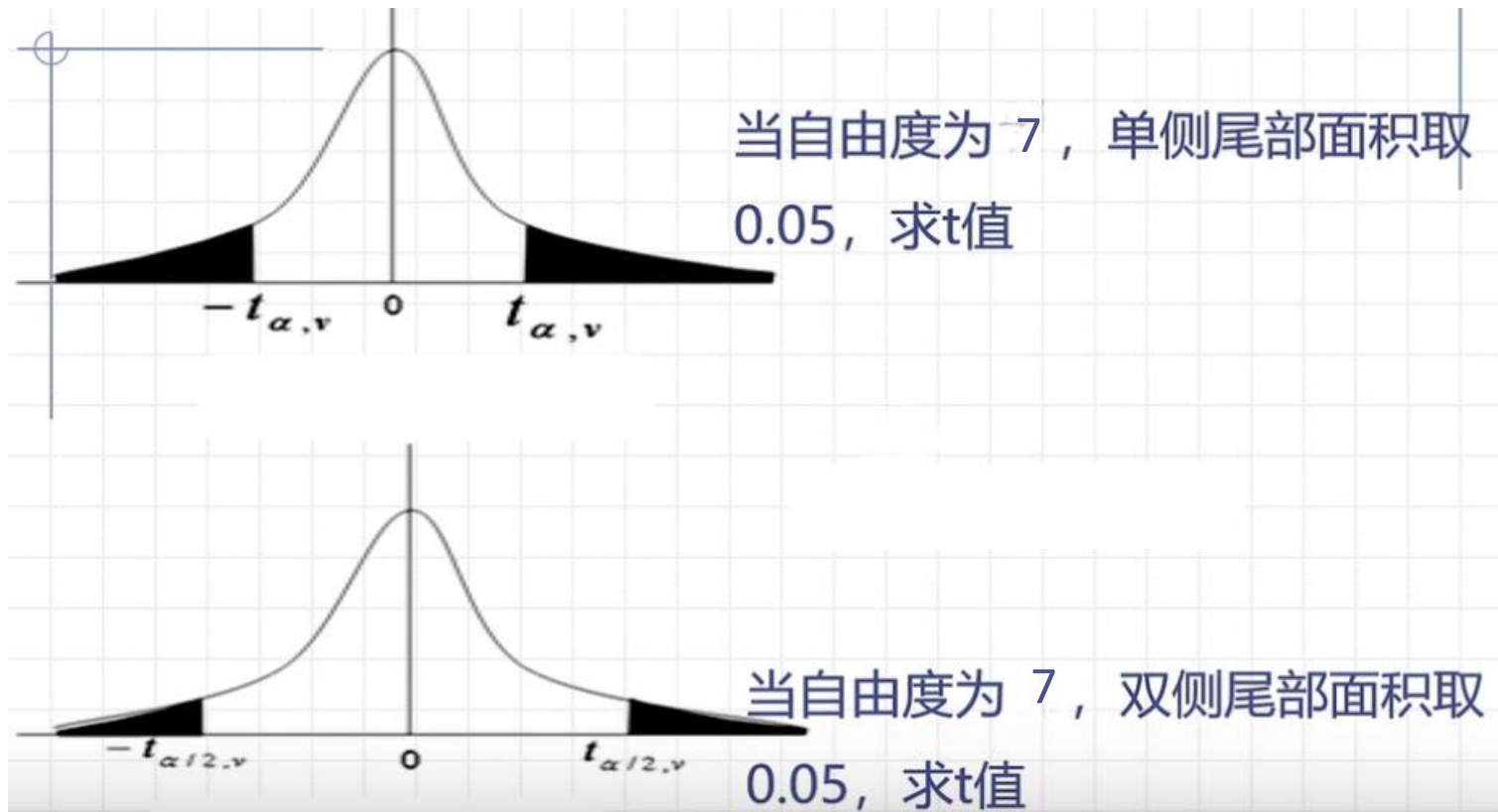


t 界值表

附表2 t界值表，给出了不同自由度情况下，单侧概率和双侧概率的t界值。

单侧	$\alpha=0.10$	0.05	0.025	0.01	0.005
双侧	$\alpha=0.20$	0.10	0.05	0.02	0.01
$V=1$	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
200	1.286	1.653	1.972	2.345	2.601
∞	1.282	1.645	1.960	2.326	2.576





$$P(-t_{\alpha/2, v} < \frac{\bar{X} - \mu}{S / \sqrt{n}} < t_{\alpha/2, v}) = 1 - \alpha$$



2. 可信区间的计算

- 计算可信区间的原理与 σ 已知情况完全相同，仅仅是两侧概率的界值有些差别。按t分布规律， $100\% \times (1 - \alpha)$ 可信区间可表示为：

$$P(-t_{\alpha/2, v} < \frac{\bar{X} - \mu}{S / \sqrt{n}} < t_{\alpha/2, v}) = 1 - \alpha$$

转化为：

$$(\bar{X} - t_{\alpha/2, v} S_{\bar{X}}, \bar{X} + t_{\alpha/2, v} S_{\bar{X}})$$





➤ 在大样本情况 ($n > 50$) 下, 可以用 $z_{\alpha/2}$ 替换 $t_{\alpha/2, v}$ 。

中心极限定理

$$(\bar{X} - z_{\alpha/2} S_{\bar{X}}, \bar{X} + z_{\alpha/2} S_{\bar{X}})$$

$$(\bar{X} - 1.96 S_{\bar{X}}, \bar{X} + 1.96 S_{\bar{X}}) \quad \sigma \text{ 未知}$$

$$(\bar{X} - 1.96 \sigma_{\bar{X}}, \bar{X} + 1.96 \sigma_{\bar{X}}) \quad \sigma \text{ 已知}$$



可信区间的计算

可信区间为 $1-\alpha$ 的计算

σ 已知

$$(\bar{X} - z_{\alpha/2} \sigma_{\bar{X}}, \bar{X} + z_{\alpha/2} \sigma_{\bar{X}})$$

σ 未知

$$(\bar{X} - t_{\alpha/2, \nu} S_{\bar{X}}, \bar{X} + t_{\alpha/2, \nu} S_{\bar{X}})$$

在大样本情况 ($n > 50$) 下, 可以用 $z_{\alpha/2}$ 替换。

$$(\bar{X} - z_{\alpha/2} S_{\bar{X}}, \bar{X} + z_{\alpha/2} S_{\bar{X}})$$





- **例6-2** 某医生测得25名动脉粥样硬化患者血浆纤维蛋白原含量的均数为3.32 g/L，标准差为0.57 g/L，试计算该种病人血浆纤维蛋白原含量总体均数的95%可信区间。

σ 未知

$$(\bar{X} - t_{\alpha/2, \nu} S_{\bar{X}}, \bar{X} + t_{\alpha/2, \nu} S_{\bar{X}})$$

$$\text{下限: } \bar{X} - t_{\alpha/2, \nu} S_{\bar{X}} = 3.32 - 2.064 \times 0.57 / \sqrt{25} = 3.08 \text{ (g/L)}$$

$$\text{上限: } \bar{X} + t_{\alpha/2, \nu} S_{\bar{X}} = 3.32 + 2.064 \times 0.57 / \sqrt{25} = 3.56 \text{ (g/L)}$$

动脉粥样硬化患者血浆纤维蛋白原含量总体均数的95%可信区间为：3.08~3.56g/L





➤ **例6-3** 试计算例6-1中该地成年男子红细胞总体均数的95%可信区间。

本例属于大样本，可采用正态近似的方法计算可信区间($\alpha = 0.05$)，

则95%可信区间为：

$$\sigma \text{ 未知, 样本量大} \quad (\bar{X} - z_{\alpha/2} S_{\bar{X}}, \bar{X} + z_{\alpha/2} S_{\bar{X}})$$

$$\bar{X} = 4.77 \quad S = 0.38 \quad n = 140$$

$$\text{下限: } \bar{X} - z_{\alpha/2} S_{\bar{X}} = 4.77 - 1.96 \times 0.38 / \sqrt{140} = 4.71$$

$$\text{上限: } \bar{X} + z_{\alpha/2} S_{\bar{X}} = 4.77 + 1.96 \times 0.38 / \sqrt{140} = 4.83$$

估计该地成年男子红细胞总体均数的95%可信区间为 $4.71 \sim 4.83 \times 10^{12} / \text{L}$ 。



3. 单侧可信区间

- 前面涉及的都是双侧可信区间。但有些情况下，我们所关心的仅仅是单侧的可信限。单侧可信区间与双侧可信区间的计算公式基本相同，只需将公式中的抽样分布的双侧界值换成单侧界值，同时只取下限或上限。

$$CI_{\text{上限}} = \bar{X} + t_{\alpha, \nu} S_{\bar{X}}$$
$$CI_{\text{下限}} = \bar{X} - t_{\alpha, \nu} S_{\bar{X}}$$





对轻度原发性高血压进行治疗，一般病人接受标准治疗后可以使舒张压降低10mmHg，现39名患者接受新方法治疗舒张压平均降低12.7mmHg，治疗前后血压变化标准差为5.6mmHg，试问新方法是否优于标准疗法？
本例 $n=100$ ， $v=38$ ，若是可信区间下限高于10mmHg，则有理由认为新方法优于标准疗法。

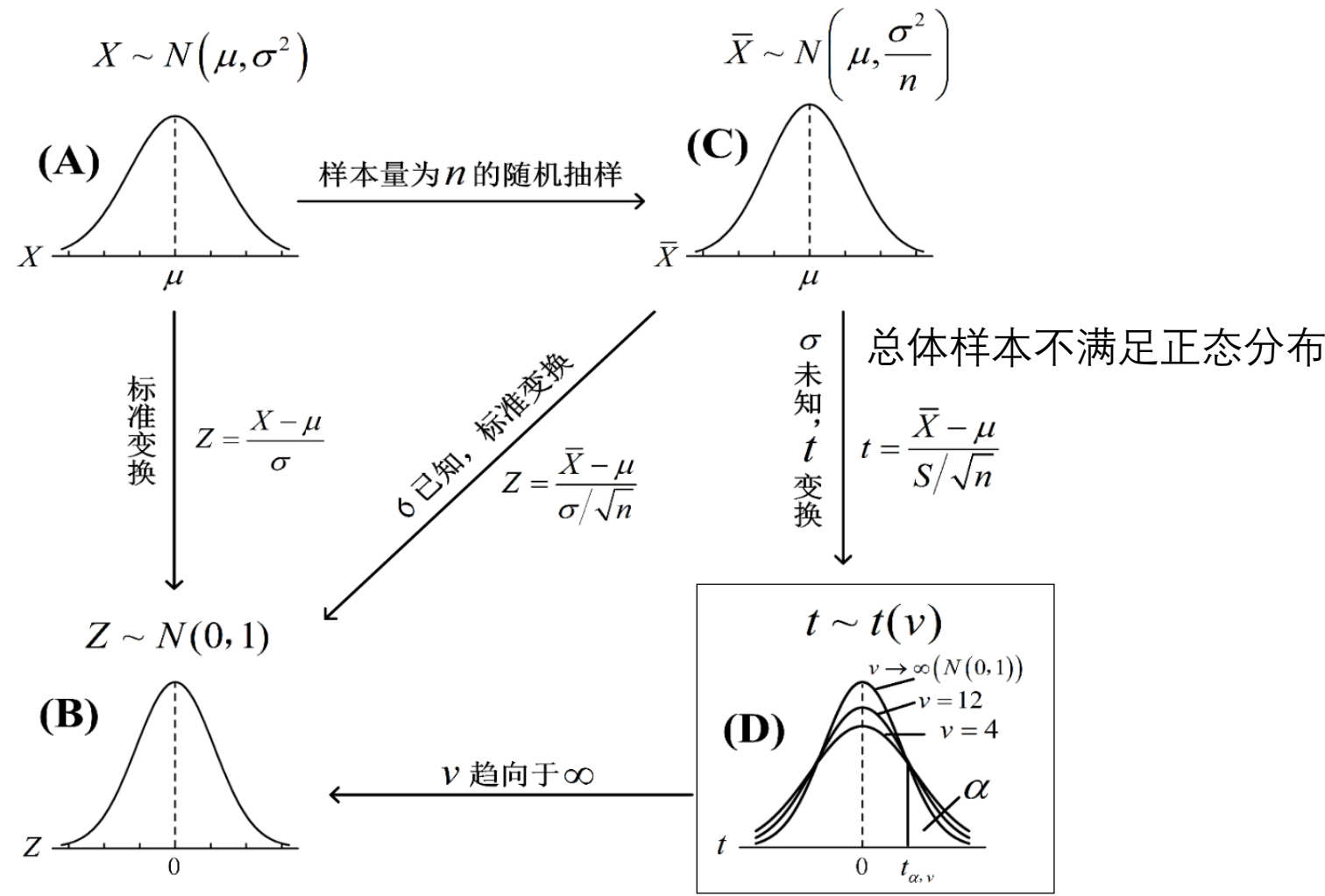
$$\bar{X} - t_{\alpha, v} S_{\bar{X}}$$

$$12.7 - 1.686 \times \frac{5.6}{\sqrt{39}} = 11.19 \text{ mmHg}$$

有95%的把握推断新方法平均降压至少11.19mmHg，优于标准治疗方法。



总结





四、两总体均数差值的区间估计*

- 实际中，有时需要计算两个总体均数差值的可信区间，例如通过计算两种降压药物平均降压的差值比较两种药物疗效的差别，其双侧 $100\% \times (1 - \alpha)$ 可信区间的计算公式为：

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2, \nu} S_{\bar{X}_1 - \bar{X}_2}$$

其中， $\nu = n_1 + n_2 - 2$ 为自由度， $S_{\bar{X}_1 - \bar{X}_2}$ 为两样本均数之差的标准误。





当两总体方差相同时,

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{S_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

其中, S_c^2 为两样本的合并方差。当两样本的样本含量均较大时, 上述计算可信区间中的 $t_{\alpha/2, v}$ 可用相应的 $z_{\alpha/2}$ 代替, 而且无论两总体的方差是否相同, 均有

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$





例题

- **例6-4** 评价复方缬沙坦胶囊与缬沙坦胶囊对照治疗轻中度高血压的有效性，将102名患者随机分为两组，其中试验组和对照组分别为54例和48例。经六周治疗后测量收缩压，试验组平均下降15.77mmHg，标准差为13.17mmHg；对照组平均下降9.53mmHg，标准差为13.55mmHg。试估计两组收缩压平均下降差值的95%可信区间。

由公式计算：

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2, \nu} S_{\bar{X}_1 - \bar{X}_2}$$

$$\bar{X}_1 = 15.77 \quad \bar{X}_2 = 9.53$$

$$S_1 = 13.17 \quad S_2 = 13.55$$





$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{S_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S_c^2 = \frac{(54 - 1) \times 13.17^2 + (48 - 1) \times 13.55^2}{54 + 48 - 2} = 178.221$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{178.221 \times \left(\frac{1}{54} + \frac{1}{48} \right)} = 2.6483$$

下限: $(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2, \nu} S_{\bar{X}_1 - \bar{X}_2} = (15.77 - 9.53) - 1.984 \times 2.6483 = 0.99$

上限: $(\bar{X}_1 - \bar{X}_2) + t_{\alpha/2, \nu} S_{\bar{X}_1 - \bar{X}_2} = (15.77 - 9.53) + 1.984 \times 2.6483 = 11.49$





五、总体率的区间估计

1. 小样本率的区间估计

利用二项分布可估计其总体率 $1-\alpha$ 可信区间, α 一般取0.05。对于 $n \leq 50$ 时, 可直接查附表6百分率的可信区间表得到其总体率的可信区间。

例6-5 2003年4—6月某医院重症监护病房收治重症SARS患者38人, 其中死亡14人, 求SARS病死率的95%可信区间。

查附表6, 在 $n = 38$ 和 $X = 14$ 的纵横交叉处的数值上行22~54 (95%置信区间), 下行18~59 (99%置信区间), 即SARS病死率95%的可信区间为22% ~ 54%。





2. 大样本率的区间估计

- 当 n 较大, p 和 $1-p$ 均不太小, 如 np 和 $n(1-p)$ 均大于5时, 可利用样本率 p 近似服从正态分布的原理来估计总体率的 $1-\alpha$ 可信区间, 计算公式为:

$$p \pm z_{\alpha/2} S_p$$

其中, $S_p = \sqrt{p(1-p)/n}$, 当 $\alpha=0.05$, $z_{0.05/2} = 1.96$





例6-6 某区疾病预防控制中心2012年对该乡镇250名小学生进行贫血的检测，结果发现有86名贫血者，检出率为34.40%，求贫血检出率95%的可信区间。

本例 n 较大，且 $np=86$ ， $n(1-p)=164$ 均大于5，用上式计算总体率95%的可信区间得：

$$p \pm z_{\alpha/2} S_p = p \pm z_{0.05/2} \sqrt{\frac{p(1-p)}{n}} = 0.3440 \pm 1.96 \sqrt{\frac{0.3440(1-0.3440)}{250}} = (0.2851, 0.4029)$$

该乡镇小学生贫血检出率95%的可信区间为28.51%-40.29%。





六、两总体率差值的区间估计*

- 设两样本率分别为 p_1 和 p_2 , 当 n_1 与 n_2 均较大, 且 $p_1, 1-p_1, p_2$ 及 $1-p_2$ 均不太小, 如 $n_1 p_1, n_1(1-p_1), n_2 p_2, n_2(1-p_2)$ 均大于5时, 可采用正态近似法对两总体率的差值进行可信区间估计, 其计算公式为:

$$(p_1 - p_2) \pm z_{\alpha/2} S_{p_1 - p_2}, \text{ 其中 } S_{p_1 - p_2} = \sqrt{p_c(1-p_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \quad p_c = \frac{X_1 + X_2}{n_1 + n_2}$$





例6-7 某医院口腔科医生用极固宁治疗牙本质过敏症，以双氟涂料作对照，进行了1年的追踪观察，结果见表，试估计两组有效率差别95%的可信区间。

治疗牙本质过敏症两组有效率的比较			
组别	总牙数	有效数	有效率(%)
试验组	77 (n_1)	61 (x_1)	79.22
对照组	69 (n_2)	38 (x_2)	55.07
合计	146	99	67.81

本例：

$$p_c = \frac{X_1 + X_2}{n_1 + n_2} = \frac{61 + 38}{77 + 69} = 0.6781$$
$$S_{p_1 - p_2} = \sqrt{p_c(1 - p_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{0.6781 \times (1 - 0.6781) \times \left(\frac{1}{77} + \frac{1}{69} \right)} = 0.0775$$

两组总体率差别95%的可信区间为：

$$(p_1 - p_2) \pm z_{\alpha/2} S_{p_1 - p_2} = (0.7922 - 0.5507) \pm 1.96 \times 0.0775 = (0.0896, 0.3934)$$



区别	均数可信区间	医学参考值
意义	按一定的置信区间 $(1-\alpha)$ 估计的总体均数所在的区间范围	大多是正常人的某项解剖、生理、生化指标的波动
计算公式	1.总体 σ 已知 2.总体 σ 未知 ($n \geq 50$) 3.总体 σ 未知 ($n < 50$)	1.正态分布 (单侧或者双侧) 2.偏态分布 (上下限)
用途	用于总体均数的估计或假设检验	观察对象的某指标正常与否, 为临床提供参考





第二节

假 设 检 验



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE



一、基本原理

表示两组数据之间的比较

- 假设检验 (hypothesis test) 亦称显著性检验 (significant test), 是统计推断的另一重要内容, 其目的是比较总体参数之间有无差别, 或总体分布是否相同。
- 假设检验的基本思想是, 首先对所需要比较的总体提出一个无差别的假设, 然后通过样本数据去推断是否拒绝这一假设。
- 做统计推断时, 应综合参数估计和假设检验两个方面来下结论。

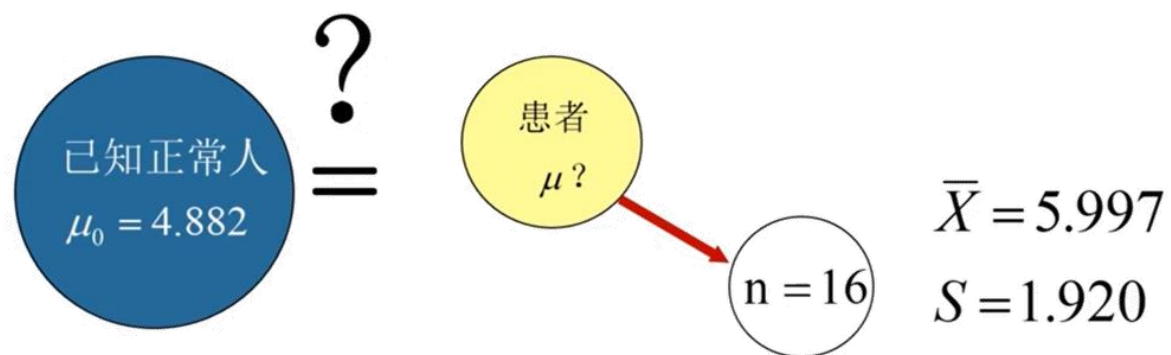




例: 已知一般无肝肾疾患的正常人群尿素氮均值为4.882mmol/L, 16名脂肪肝患者的尿素氮平均值为5.997 mmol/L, 标准差为1.920mmol/L。问脂肪肝患者尿素氮水平是否与正常人相同?

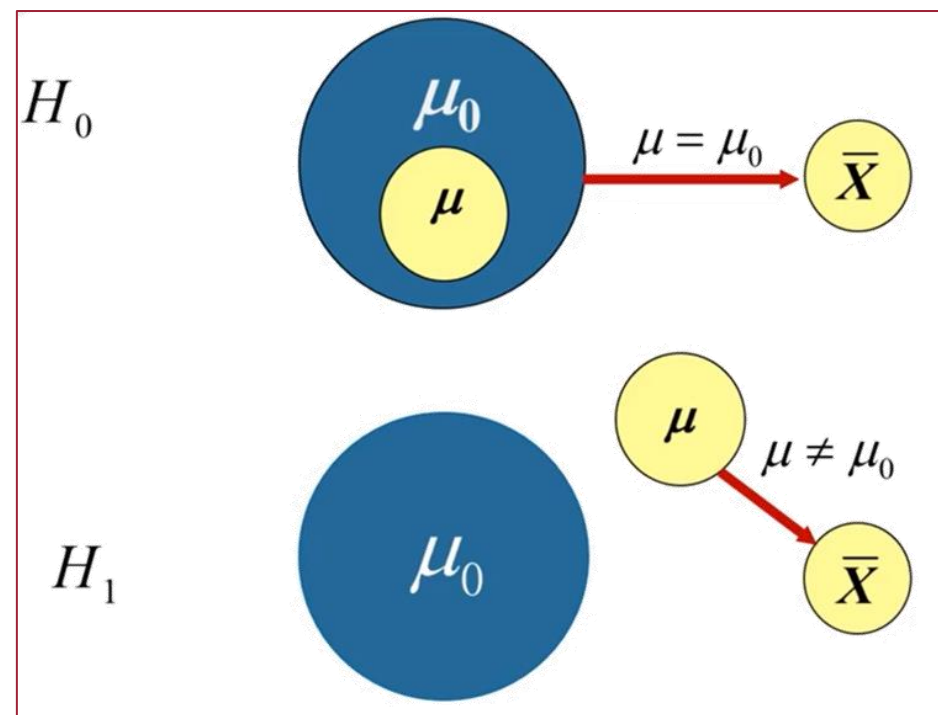
$$\text{已知: } \mu_0 = 4.882 \quad \bar{X} = 5.997 \quad S = 1.920 \quad n = 16$$





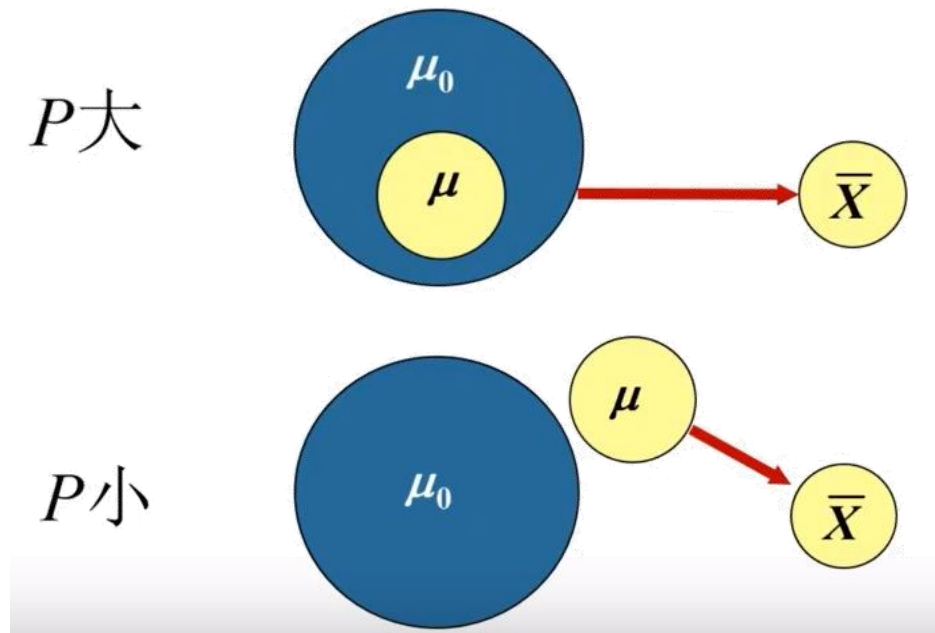
①同一总体，即 $\mu = \mu_0$ 但有抽样误差存在；

②非同一总体，即 $\mu \neq \mu_0$ 存在本质上的差别，同时有抽样误差存在。



反证法思想

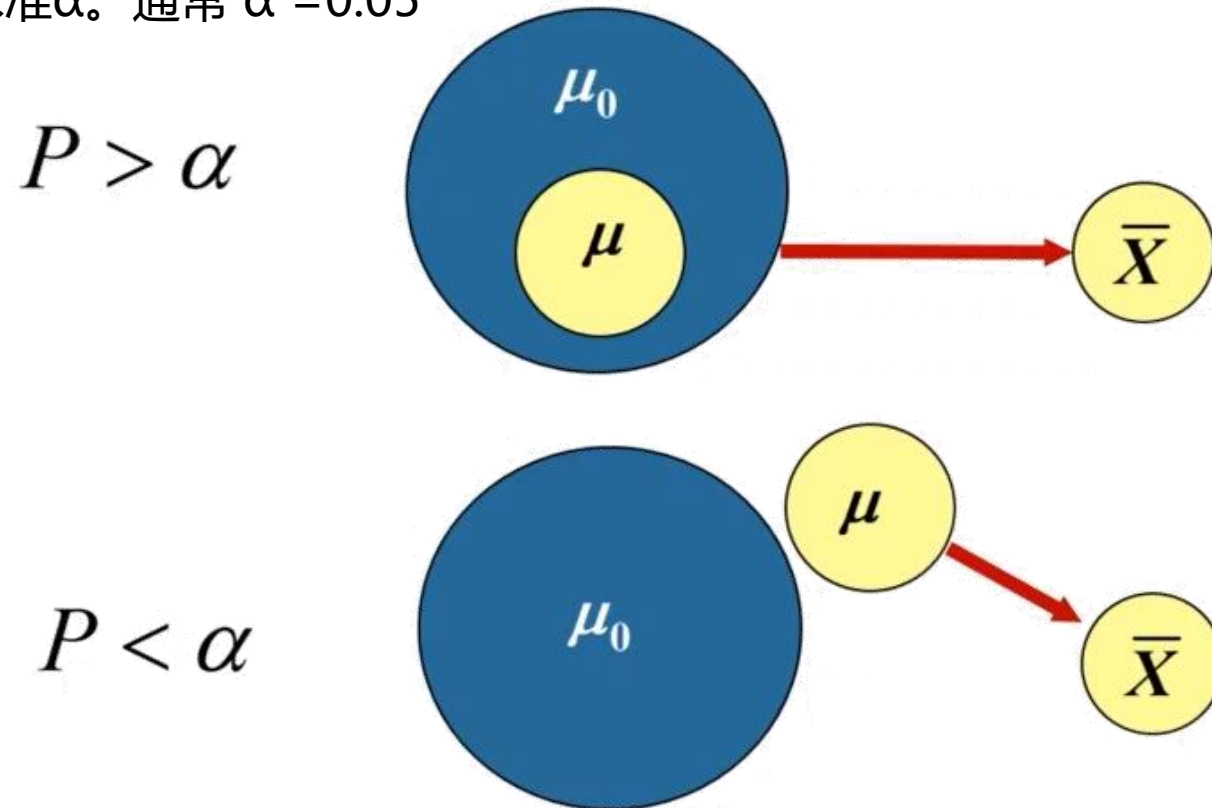
首先假设 $\mu = \mu_0$ (H_0 假设), 用适当的统计方法确定从 μ_0 总体中抽得当前样本或者更极端样本的可能性的 P 值大小, 如果可能性大, 则不能拒绝 H_0 ; 如果可能性小, 则认为 H_0 不成立, 拒绝它。



小概率思想

小概率事件在一次随机抽样中基本上不会发生。

概率小于多少算小概率是相对的，在进行统计分析时要事先规定，即检验水准 α 。通常 $\alpha = 0.05$

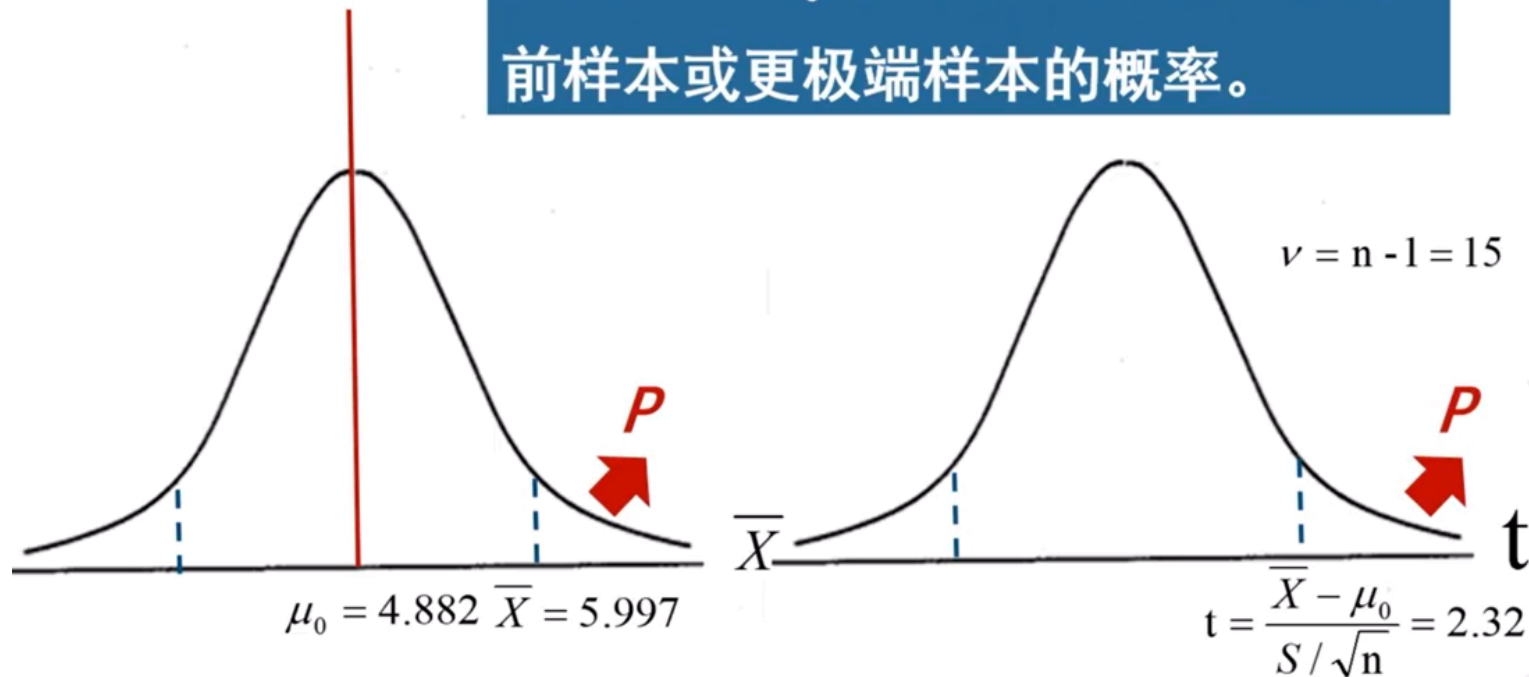




例: 已知一般无肝肾疾患的正常人群尿素氮均值为4.882mmol/L, 16名脂肪肝患者的尿素氮平均值为5.997 mmol/L, 标准差为1.920mmol/L。问脂肪肝患者尿素氮水平是否与正常人相同?

$$\text{已知: } \mu_0 = 4.882 \quad \bar{X} = 5.997 \quad S = 1.920 \quad n = 16$$

P 值: 从 H_0 规定的总体中抽得当前样本或更极端样本的概率。





如果治疗前后血清甘油三酯没有改善，则前后变化差值的总体均数应为 $\mu_d = 0$

但现在得到的样本差值均数 $\bar{d} = 1.38$ ，其差别可能是两种原因引起：

- 药物作用
- 抽样误差

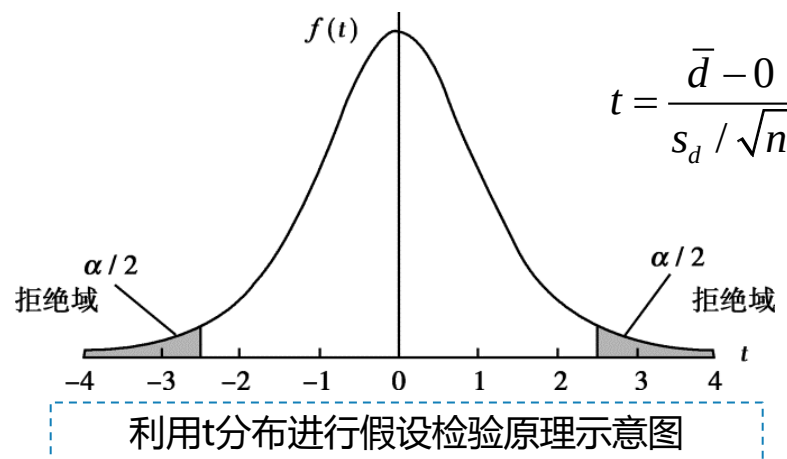
即使治疗前后血清甘油三酯没有改善，由于个体变异和抽样误差， $\bar{d}$ 也不会恰好等于0，但当 $\bar{d} - 0$ 较大时，血清甘油三酯的改变难以完全用抽样误差来解释，从而认为药物具有作用。

问题： $|\bar{d} - 0|$ 究竟多大能够下“有效”的结论？



说明

- 在 $H_0: \mu_d = 0$ 成立的条件下, $t = (\bar{d} - 0) / (S_d / \sqrt{n})$ 服从 t 分布 (S_d 为差值的标准差)。因此, 可以利用现有样本计算 t 统计量, 然后根据 t 分布得到相应的概率 P 进行判断, 若 $P < \alpha$, 则倾向于拒绝假设 H_0 , 认为治疗前后有差别。





假设检验的基本思想

假设检验的方法是首先**假设**样本对应的总体参数与某个已知总体参数相同，然后根据某样本统计量的抽样分布规律，**计算**出从假设的总体中抽得该样本或更极端样本的可能性的**大小**，以**判断**样本信息是否支持这种假设，并对假设做出取舍抉择。





二、基本步骤

(一) 建立假设、确定检验水准

- 假设检验中，包括无效假设（null hypothesis 零假设）和备择假设（alternative hypothesis）两种假设。

1. 无效假设（无差别假设）

无效假设符号为 H_0 ，指需要检验的假设，如治疗前后血清甘油三酯没有差别，即 $H_0: \mu_d = 0$ ；这一假设通常与我们要验证的结论相反。





2. 备择假设

- 备择假设符号为 H_1 ，是在 H_0 成立证据不足的情况下而被接受的假设，如拒绝治疗前后血清甘油三酯相同的假设，可表示为 $H_1: \mu_d \neq 0$ 。
- 备择假设有**双侧**和**单侧**两种情况。双侧检验指无论是正方向还是负方向的误差，若显著地超出界值则拒绝 H_0 ， $H_1: \mu_d \neq 0$ 即为双侧检验。单侧检验指仅在正方向或负方向误差超出规定的界值时则拒绝 H_0 ，如治疗后血清甘油三酯下降的假设可表示为 $H_1: \mu_d > 0$ 。
- 双侧检验和单侧检验应如何选择，需根据研究目的和专业知识而定。





3. 检验水准

- 建立检验假设的同时, 还必须给出检验水准。检验水准亦称显著性水准 (significant level), 用 α 表示, 是预先规定的拒绝域的概率值, 实际中一般取 $\alpha = 0.05$ 或 $\alpha = 0.01$ 。

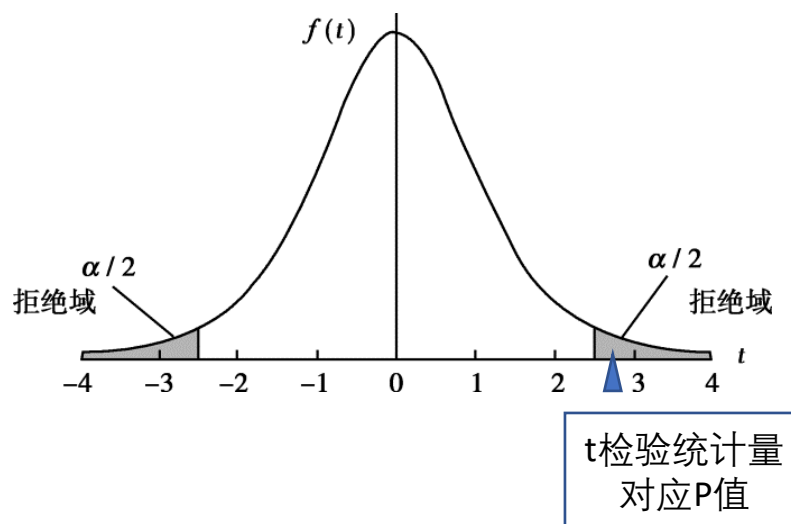
(二) 选择检验方法、计算检验统计量

- 根据资料类型、研究设计方案和统计推断的目的, 选择适当的检验方法和计算公式。许多假设检验方法是以检验统计量来命名的, 如 t 检验、 z 检验、 F 检验和 χ^2 检验等。



(三) 根据 P 值做出统计推断

- 查表得到检验用的临界值，然后将算得的统计量与拒绝域的临界值作比较，。
- 如对双侧 t 检验 $|t| \geq t_{\alpha/2, \nu}$ ，则 $P \leq \alpha$ ，按 α 检验水准拒绝 H_0 ，接受 H_1 ；若 $P > \alpha$ ，则不能拒绝 H_0 。

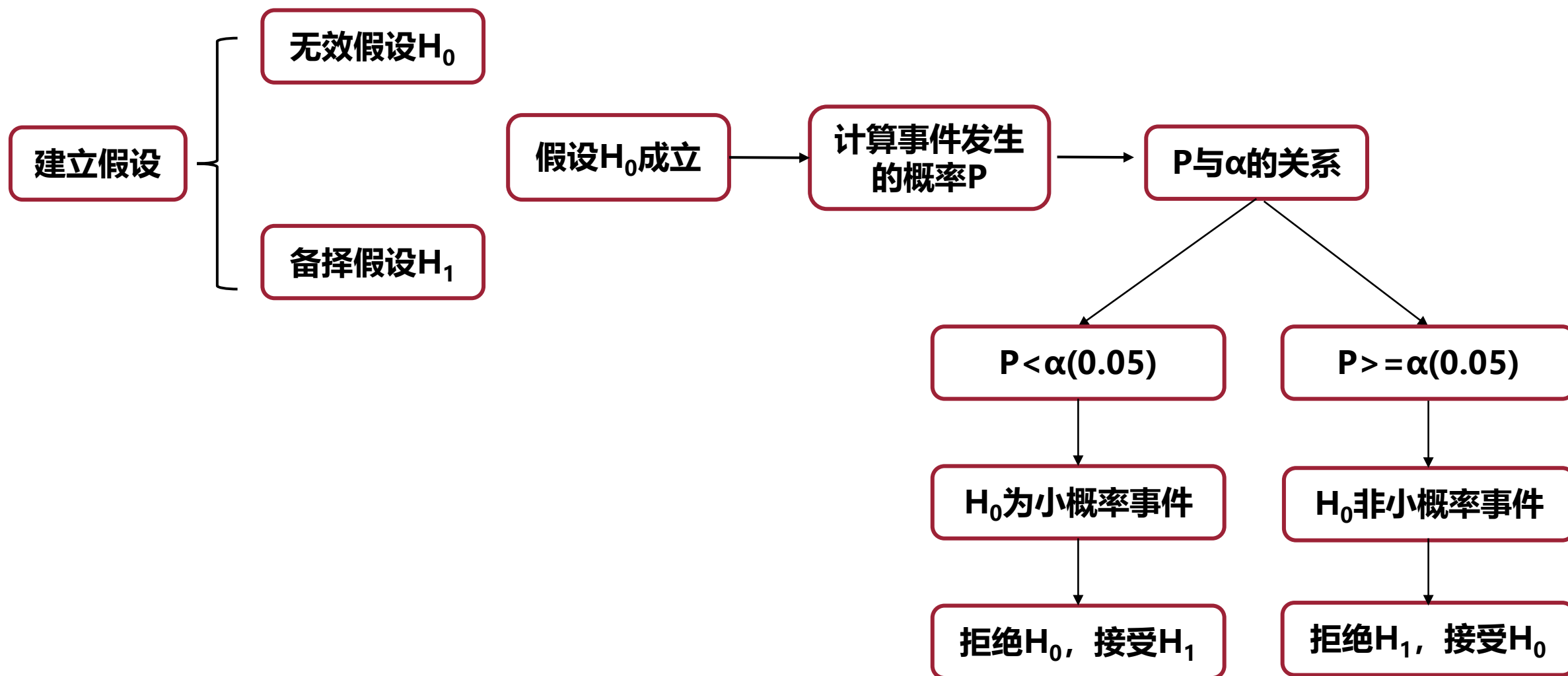


- 需要注意：不拒绝 H_0 不等于支持 H_0 成立，仅表示现有样本信息不足以拒绝 H_0 。



- **P 值是假设检验下结论的主要依据**，其含义是指在无效假设 H_0 成立的条件下，观察到的比现有样本更极端情况的概率，也就是拒绝 H_0 的风险，如果此风险足够小，就可以安全地拒绝原假设。因此， P 值越小越有理由拒绝 H_0 ，认为总体之间有差别的统计学证据越充分。由于统计软件的普及，下结论时，最好给出 p 值，以便对检验结果的可靠性做出准确的判断。





第三节

假设检验中的两类错误



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE



(一) I 类错误和II类错误

假设检验实际上是一种风险决策，即根据**现有样本信息**来定性地判断无效假设 H_0 是否可以被拒绝。

这种基于**概率**的决策，不论做出哪一种推断结论，都有可能发生错误。





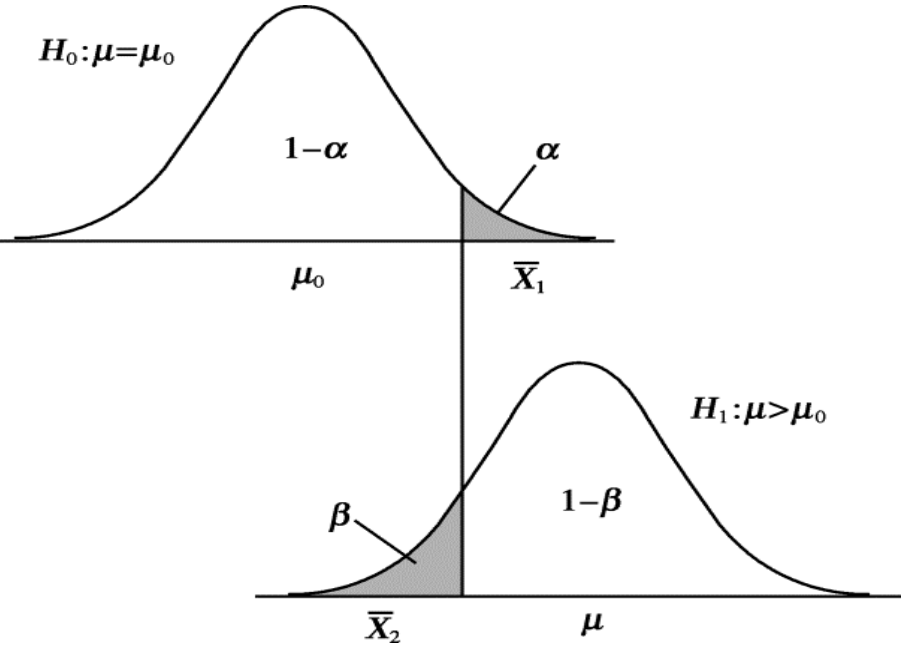
1. I 类错误

- 如果真实情况与检验假设 H_0 一致, 仅仅由于抽样误差, 使得检验统计量的值落到拒绝域, 导致推断结论错误, 对此称为第 I 类错误 (type I error), 用 α 表示。 H_0 真, 拒绝 H_0 , 假阳性。

2. II类错误

- 如果真实情况与检验假设 H_0 不一致, 检验统计量的值却落到了接受域, 则导致了另一种推断错误, 即第II类错误 (type II error), 用 β 表示。 H_1 真, 接受 H_0 , 假阴性。





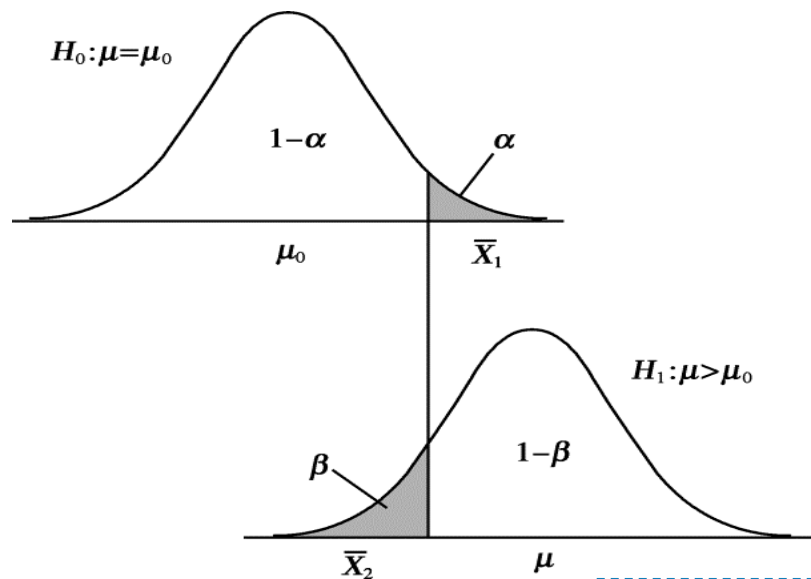
统计推断的两类错误及其概率

真实情况	统计推断--假设检验结论	
	拒绝 H_0	不拒绝 H_0
H_0 成立, 无差异	I 类错误(α) (假阳性)	推断正确 ($1-\alpha$)
H_1 成立, 有差异	推断正确($1-\beta$)	II 类错误(β) (假阴性)

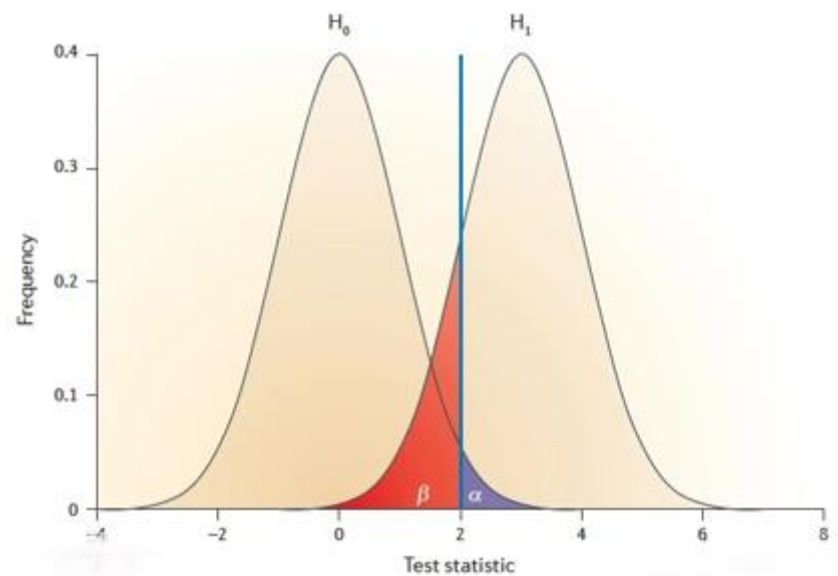


3. I 类错误与II类错误的关系

- 当样本量一定时, 给定的检验水准 α 值越小, 出现II类错误的概率 β 越大, 反之亦然。
- 若要同时减小 α 和 β , 可以通过增加样本量的方法实现。

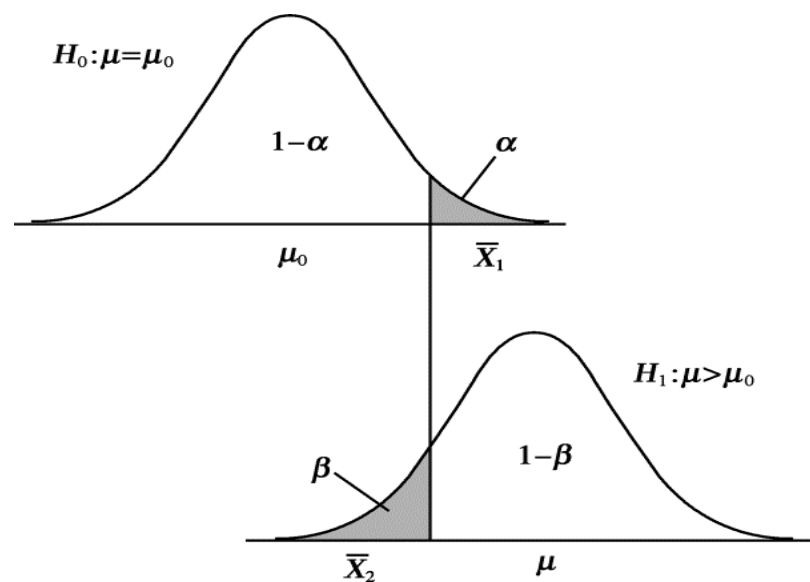


I 类错误和II类错误示意图



检验效能

指当不同总体间确实有差别时，按规定的检验水准 α ，能发现其差别的概率，其值为 $1-\beta$ 。





总结

- 假设检验是依据样本提供的有限信息对总体作出推断的统计学方法，是在对研究总体两个对立假设作抉择的过程。

- 假设检验的过程：

确定两个假设→选择并计算统计量→与 α 值对应的统计量比较→结论

- 当样本量一定时，给定的检验水准 α 值越小，出现Ⅱ类错误的概率 β 越大，反之亦然。



第四节

假设检验与区间估计的关系



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE



- 假设检验与区间估计是统计推断的两个方面，其中前者是对两总体关系的一个定性决策，而后者是对参数关系的定量概率描述。
- 对同一数据，两者可得到同样的结论。

例如，前述治疗前后血清甘油三酯差值的问题，如差值总体均数 $1-\alpha$ 可信区间不包含0，则按照相同 α 检验水准也会得到拒绝 H_0 的推断结论。

- 可信区间与假设检验结合起来，可以提供更为全面、完整的信息



本章小结

1. 反映抽样误差大小的指标是标准误。均数标准误的估计值为 $S_{\bar{x}} = S/\sqrt{n}$; 率的标准误估计值为 $S_p = \sqrt{p(1-p)/n}$ 。
2. **参数估计**指由样本统计量估计总体参数，是统计推断的重要内容之一，常用的方法有**点估计**和**区间估计**。区间估计是指按预先给定的可信度，计算出一个区间，使它能够包含未知的总体参数。区间越窄说明估计的准确度越高。



本章小结

3. **假设检验**是统计推断的另一重要内容，其目的是定性比较总体参数之间有无差别。基本思想是，首先对所需要比较的总体提出一个无差别的假设，然后通过样本数据去推断是否拒绝这一假设，得出结论的依据是 P 值。

4. 假设检验与区间估计是统计推断的两个方面，可信区间用于推断总体均数的具体数量范围，而假设检验用于推断总体均数间是否不同。两者既有区别，又有联系。

本章小结

5. 拒绝实际上成立的 H_0 ，所犯的误差称I类误差，其概率的最大值为 α ；不拒绝实际上不成立的检验假设所犯的误差称为II类误差。当样本含量一定时，给定检验水准 α 值越小，出现II类错误的概率越大，反之亦然。若要同时减少两类误差，可以通过增加样本量的方法实现。





课后习题二-1

样本含量为 450，属于大样本，可采用正态近似的方法计算可信区间。

$$\bar{X} = 101.4, S = 1.5, n = 450$$

$$S_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{1.5}{\sqrt{450}} = 0.07$$





95%可信区间为:

$$\text{下限: } \bar{X} - u_{\alpha/2} \times S_{\bar{x}} = 101.4 - 1.96 \times 0.07 = 101.26(\text{g/L})$$

$$\text{上限: } \bar{X} + u_{\alpha/2} \times S_{\bar{x}} = 101.4 + 1.96 \times 0.07 = 101.54(\text{g/L})$$

即该地小学生血红蛋白的 总体均数的 95%可信区间为 101.26~101.54g/L。





课后习题二-2

根据已知条件, 可知, $n_1=102$, $n_2=104$, $p_1=94.4\%$, $p_2=91.26\%$

$$p_c = \frac{X_1 + X_2}{n_1 + n_2} = \frac{102 \times 0.944 + 104 \times 0.9126}{102 + 104} = 92.8\%$$

$$S_{p_1-p_2} = \sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0.928(1-0.928)\left(\frac{1}{102} + \frac{1}{104}\right)} = 0.036$$





两组有效率差值 95%的可信区间为:

$$(p_1 - p_2) \pm z_{\alpha/2} S_{p_1 - p_2} = (0.944 - 0.9126) \pm 1.96 \times 0.036 = (-0.039, 0.102)$$

即两组治疗细菌感染有效率差值 95%的可信区间为 (-3.9%, 10.2%) 。





课后习题二-3

①均数的标准误可以用来衡量样本均数的抽样误差大小，即

$$S = 30, n = 100, S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{30}{\sqrt{100}} = 3.0$$





②样本含量为 100, 属于大样本, 可采用正态近似的方法计算可信区间。

$$\bar{X} = 207.5, S = 30, n = 100, S_{\bar{x}} = 3$$

则 95%的可信区间为:

$$\text{下限: } \bar{X} - u_{\alpha/2} \times S_{\bar{x}} = 207.5 - 1.96 \times 3 = 201.62(\text{mg/dl})$$

$$\text{上限: } \bar{X} + u_{\alpha/2} \times S_{\bar{x}} = 207.5 + 1.96 \times 3 = 213.38(\text{mg/dl})$$

故该地 100 名儿童的胆固醇平均水平的 95%可信区间为 201.62~213.38mg/dl。





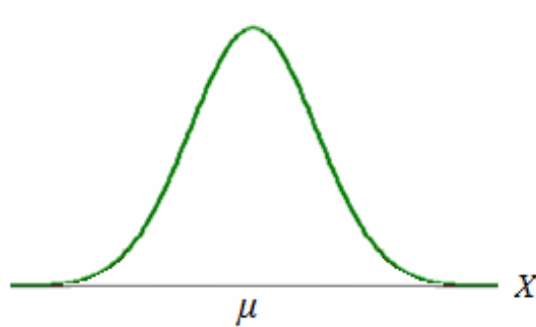
③因为 100 名曾患心脏病且胆固醇高的子代儿童的胆固醇平均水平的 95%可信区间的下限高于正常儿童的总胆固醇平均水平 175mg/dl, 提示患心脏病且胆固醇高的父辈, 其子代胆固醇水平较高, 即高胆固醇具有一定的家庭聚集性。





t检验的理论依据是假设检验原理

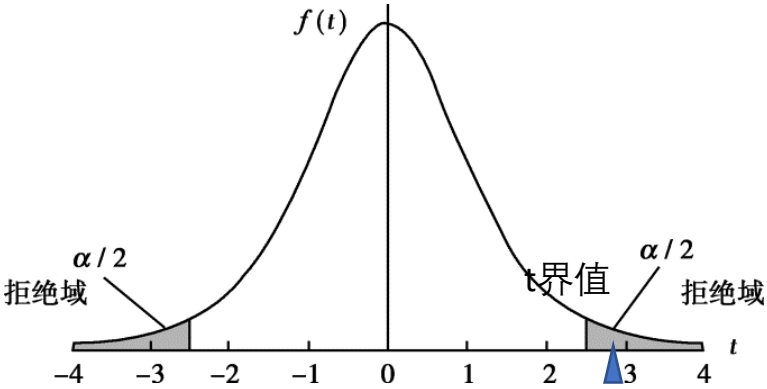
样本1代表的总体A



样本2代表的是同一总体，即同为总体A



t检验的理论依据是假设检验原理



该p值代表若两样本代表的总体为同一个时，出现现有样本均值及更极端值的概率

t检验统计量
对应P值

$$t = (\bar{X} - \mu) / (S / \sqrt{n})$$

判断样本1与样本2所代表的总体是否是同一个时，构建一个t统计量，该统计量服从左侧分布，我们划定了一个拒绝域，当样本1和样本2来自同一总体时，样本2均值的t检验统计量落在该拒绝域的概率为 α ，为犯第 I 类错误的概率，也就是假阳性的概率。

可理解为医学参考值范围

正如我们取 $\alpha=0.05$ ，对应的有一个t界值，通过比较样本2均值的t检验统计量与t界值，若 $>t$ 界值，认为落入拒绝域，拒绝零假设，认为两样本代表的总体有差别。实际上，计算的样本2均值的t检验统计量，会对应一个p值，若 $P<\alpha$ ($P<0.05$)，则拒绝零假设。在统计软件计算时，通常给出P值 ($P<0.05$)，以说明差异的显著性。



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE

谢谢观看