# 统计思维与
# 美国统计学会关于P值的声明

方积乾

中山大学公共卫生学院

**2016.5**

# 预备知识

设 $X_1, X_2, \cdots, X_n$ 独立同分布, 服从 $N(\mu, \sigma^2)$

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

则 $\bar{X}$ 服从 $N(\mu, \frac{\sigma^2}{n})$

$\frac{\bar{X} - \mu}{\sigma / n}$ 服从 $N(0,1)$

$\frac{\bar{X} - \mu}{S / n}$ 服从 ***t-distribution***

# 置信区间的统计思维

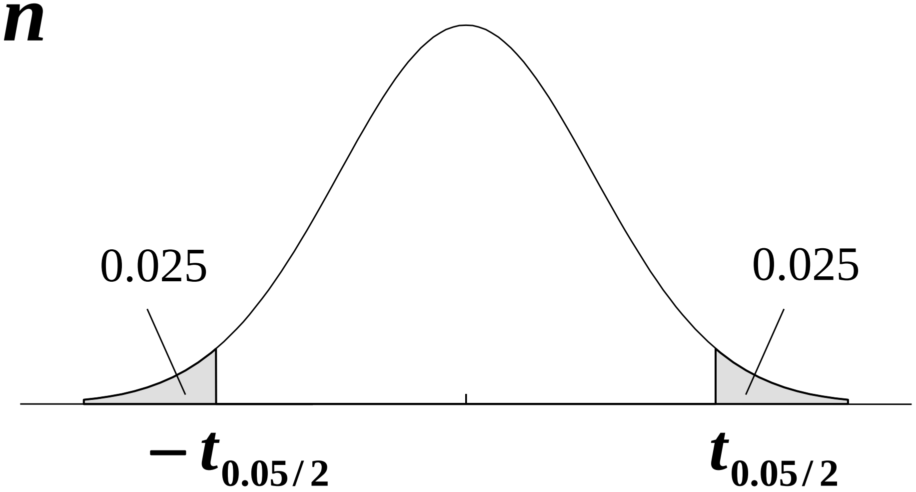
**(1)** 基于样本资料，可以估计总体性质；

**(2)** 基于样本资料估计总体性质，不可能绝对准确，只能给一个区间；

**(3)** 允许犯错误，但是保证一定的可信度。

**(4)** 放弃一定的可信度，将“无限”转化为“有限”。

$$\frac{\overline{X}-\mu}{S/n}$$ 服从 ***t-distribution***




已知 $\overline{X}, S^2$ 如何估计 $\mu$ ?

**(1)**

$$-\infty < \frac{\overline{X}-\mu}{S/n} < +\infty, \qquad -\infty < \mu < +\infty$$

**(2)** 牺牲**5%**，化无限为有限

$$-t_{0.05/2} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < +t_{0.05/2}, \quad \overline{X} - t_{0.05/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{0.05/2}\frac{S}{\sqrt{n}}$$

仅在**95%**的场合正确，故可相信的程度为**95%**！

## 表 6-1　从正态总体 $N(155.4, 5.3^2)$ 抽出的 100 份随机样本的计算结果 $(n_i = 30)$

| 样本号 (1) | 样本均数 $\overline{X}_i$ (2) | 标准误 $S_{\overline{X}}$ (3) | 95%置信区间 95%CI (4) | 样本号 (1) | 样本均数 $\overline{X}_i$ (2) | 标准误 $S_{\overline{X}}$ (3) | 95%置信区间 95%CI (4) |
|---|---|---|---|---|---|---|---|
| 1 | 156.7 | 0.91 | 154.8~158.6 | *57 | 158.2 | 0.97 | 156.2~160.2 |
| *2 | 158.1 | 0.95 | 156.2~160.1 | 58 | 154.9 | 1.06 | 152.7~157.1 |
| 3 | 155.6 | 1.16 | 153.3~158.0 | *59 | 153.4 | 0.91 | 151.5~155.3 |
| 4 | 155.2 | 1.03 | 153.1~157.3 | 60 | 156.4 | 0.98 | 154.4~158.4 |
| … | … | … | … | … | … | … | … |
| 51 | 155.7 | 0.97 | 153.7~157.7 | *96 | 152.7 | 0.75 | 151.1~154.2 |
| *52 | 153.7 | 0.80 | 152.1~155.4 | 97 | 155.1 | 0.93 | 153.2~157.0 |
| *52 | 153.7 | 0.80 | 152.1~155.4 | .. | .. | … | … |
| … | … | … | … | 100 | 156.6 | 1.16 | 154.2~159.0 |

*表示利用该样本资料求得的 95%置信区间未包含已知总体均数 155.4

# 假设检验的统计思维

(1) 实践中，面对事物的不确定性，人们往往要估算失误发生的概率；

(2) 如果某事物发生的概率很小（即不大可能发生），为了决策和行动，便**将“不大可能”当作“不可能”**。

例如，

- 行走在马路上，被汽车压死的概率很小，故人们照样上街；
- 飞机失事的概率很小，故坐飞机的人越来越多；
- 走在大楼前，恰好有人跳下，砸死行人的概率很小，故无人胆战心惊过大楼。

(3) 司法中的“无罪推断”

面对犯罪嫌疑人，人们想证实其为罪人；

在弄清事实之前，法官总是假定此人与好人无区别；

原告的任务是收集证据，推翻无罪假定；

仅当法官认为一个好人不大可能出现目前的行为时，方可质疑无罪假定，乃至推翻无罪假定；推翻无罪假定，法官才宣布：此人不同于好人。

若法官认为一个好人颇有可能出现目前的行为时，不能推翻无罪假定，只能宣布：证据不足，无罪释放；但并不意味此人就是好人。

例　已知北方农村儿童前囟门闭合月龄均值为 14.1 月。有人从东北某县抽取 36 名儿童，得前囟门闭合月龄均值为 14.3 月，标准差为 5.08 月。问该县儿童前囟门闭合月龄是否大于一般儿童闭合的月龄？

1. 建立检验假设，确定检验水准

$H_0$：$\mu$ =14.1，$H_1$：$\mu \neq$14.1（双侧）

注：相当于司法中的“无罪假设”

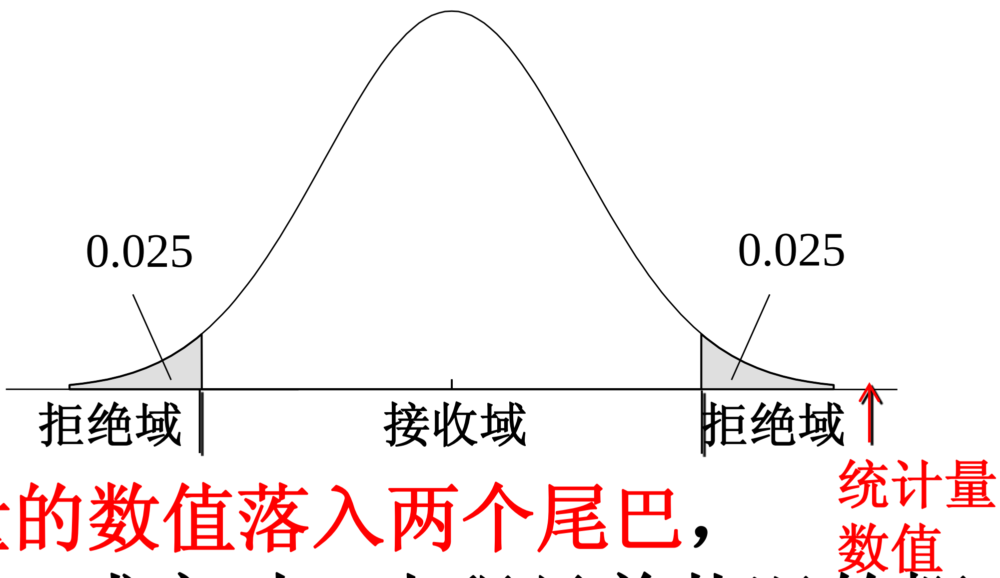
$H_0$：此人与好人无区别

$H_1$：此人不同于好人

## 2. 如果 $H_0$ 成立，出现目前状况的可能性？

利用目前样本数据计算统计量：$t = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}}$

意义：样本均数离 $\mu_0$ 的距离（用标准误作单位）

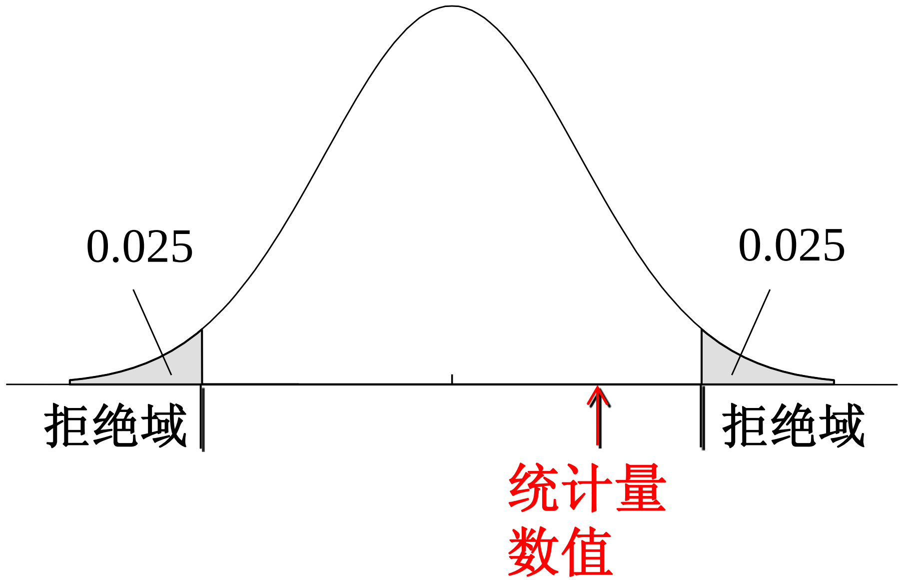



- 若统计量的数值落入两个尾巴，

(1) 表明“$H_0$ 成立时，出现目前状况的概率很小”

——不大可能出现目前状况;

(2) 把“不大可能”当作“不可能”：“$H_0$ 成立时，不可能出现目前状况”——$H_0$ 不能成立, 故拒绝！




若统计量的数值落入中间区域，

(1) 表明“ $H_0$ 成立时，出现目前状况的可能性不小”

—— 即颇有可能出现目前状况；

(2) 没有足够理由质疑 $H_0$

—— 尚不能拒绝 $H_0$ ！

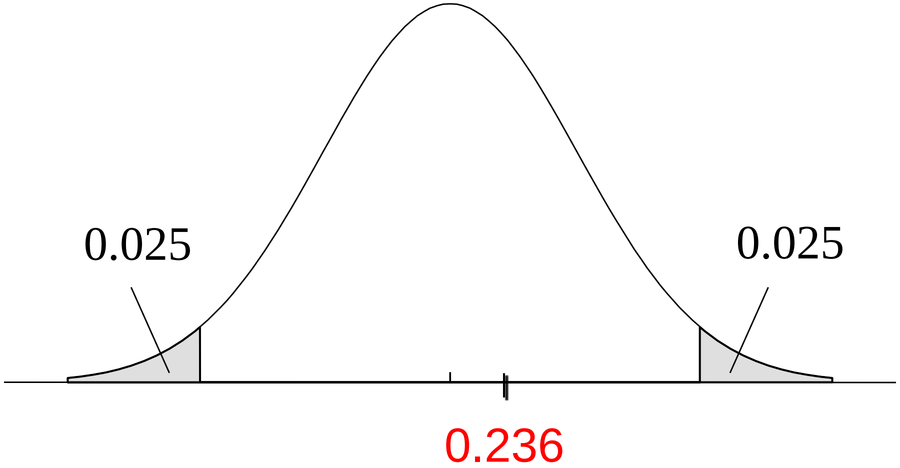
(3) 尚不能拒绝 $H_0$ , 并不意味着肯定 $H_0$

例　$H_0$：$\mu$ =**14.1**（月）成立时，

将样本数据代入，

$$t=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}=\frac{14.3-14.1}{5.08/\sqrt{36}}=0.236$$

自由度　　$\nu=n-1=36-1=35$




**(1)** 表明"$H_0$成立时，出现目前状况的可能性不小"

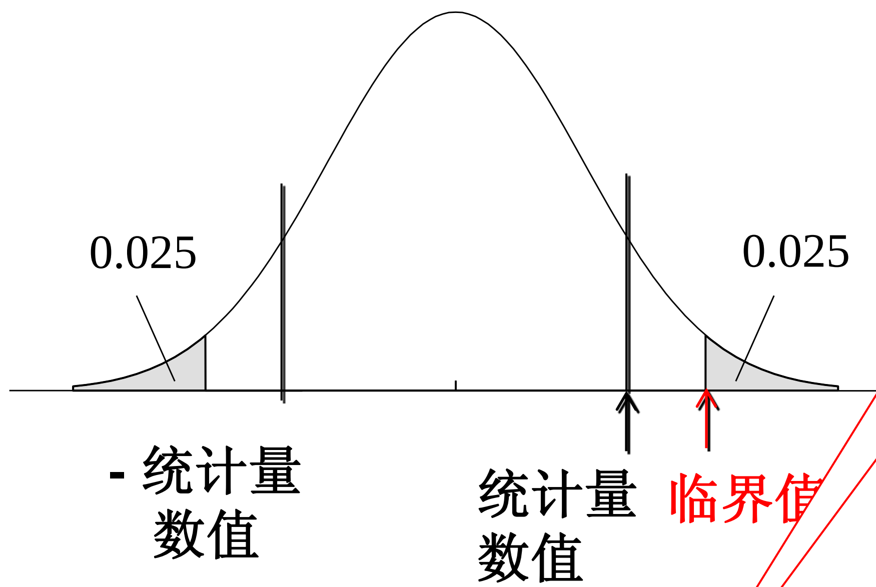
—— 即颇有可能出现目前状况;

**(2)** 没有足够理由质疑 $H_0$ —— 尚不能拒绝 $H_0$ ！

# P-值的定义

第二种说法的好处？

不需要临界值，直接描述多麽“离谱”。




两种说法：

(1) 统计量数值 <临界值

或 (2) 统计量数值以外的“尾巴”较大 (> 0.025)

单侧P-值= …… (> 0.025)

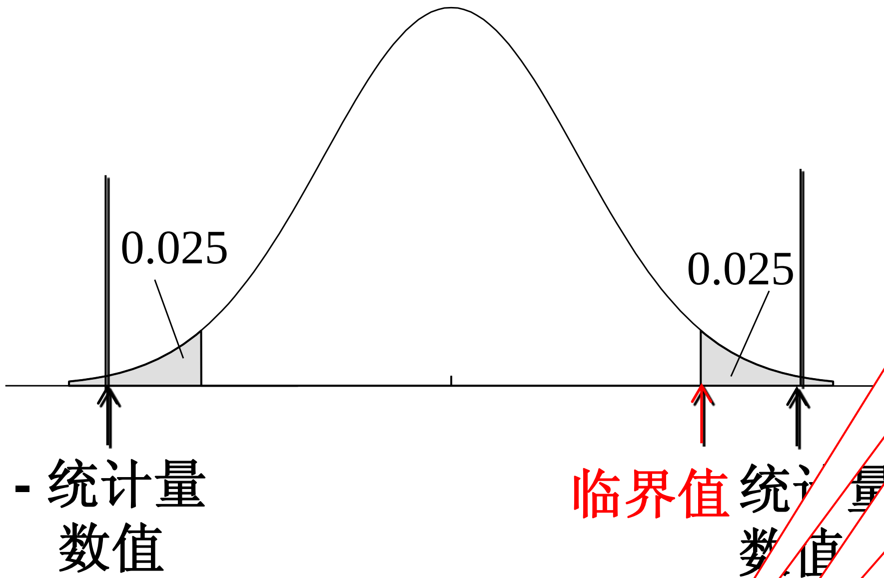
± 统计量数值以外的两个“尾巴”较大 (> 0.05)

双侧P-值= …… (> 0.05)

第二种说

不需要临界值，直接描述多麼“离谱”。

# P-值的定义




两种说法:

(1) 统计量数值 > 临界值

或 (2) 统计量数值以外的“尾巴”较小 (< 0.025)

单侧P-值= …… (< 0.025)

± 统计量数值以外的两个“尾巴”较小 (< 0.05)

双侧P-值= …… (< 0.05)

# 美国统计学会关于*P*-值的声明 (2016.1)

1. ***p*-**值可以表明数据和特定统计模型之间如何不相容。
2. ***p*-**值并不度量研究假设为真的概率, 或者数据纯系随机产生的概率。
3. 科学结论和商务或政策决定不可以仅仅基于一个***p*-**值是否通过特定的阈值。
4. 正确恰当的推断要求完整的报告和透明度。
5. ***p*-**值或统计学意义并不度量效应的大小或结果的重要性。
6. ***p*-**值本身并不对模型或假设提供一个好的度量。

# 小结

**1.** 置信区间的统计思维：

基于样本资料，可以估计总体性质，但不可能绝对准确，只能给一个区间；

如果不允许犯错误，不能给一个无意义的区间，例如，$-\infty < \mu < +\infty$ ；

如果允许少数场合犯错误，可以给出一个置信区间，例如，

$$\bar{X} - t_{0.05/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{0.05/2}\frac{S}{\sqrt{n}}$$

**2. 假设检验的统计思维：**

司法中的“无罪推断”

面对犯罪嫌疑人，法官总是假定此人与好人无区别（零假设）；原告的任务是收集证据，推翻无罪假定（对立假设）；

仅当一个好人不大可能出现目前的行为时，方可质疑无罪假定，乃至推翻无罪假定，宣布此人不同于好人。

若一个好人颇有可能出现目前的行为时，不能推翻无罪假定，只能宣布：证据不足，无罪释放；但并不意味此人就是好人。

**3.** 关于***p*-**值，务必正确地教给学生，千万不要误人子弟

（**1**）***p*-**值表明数据和零假设之间如何不相容

（**2**）***p*-**值和"统计学意义"并不度量：

研究假设为真的概率、

效应的大小、

结果的重要性、

模型或假设的好坏

（**3**）正确恰当的推断要求完整的报告和透明度

# 谢谢