第七章 t 检验

第一节 t 检验

一、单样本 t 检验（one-sample t-test）

一句总纲：
单样本 t 检验是在问：
“我这次样本均值离标准值这么远，是‘真偏离’，还是‘抽样运气’？”

---

1 什么时候用（Use case）

你有：

- 一组样本（来自某总体）
- 一个已知的“标尺均值” μ_0（标准值/理论值/大样本稳定值）
- 但总体标准差 σ **未知**（只能用样本标准差 S 代替）

你想检验：

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

---

2 t 值到底是什么（核心直觉）

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}, v = n - 1$$

把它拆成一句人话：

t = “偏离了多少” ÷ “这种偏离本来就会有多抖”

- 分子 $\bar{X} - \mu_0$：你看到的偏离（signal）
- 分母 $S/\sqrt{n}$：均值的随机波动尺度（noise / standard error）

📌 所以 t 检验本质不是比均值大小，而是比：**偏离 / 噪声**（signal-to-noise）。

【例 7-1】以往通过大规模调查已知某地新生儿平均出生体重为 3.30kg。从该地难产儿中随机抽取 35 名新生儿作为研究样本，平均出生体重为 3.42kg，标准差为 0.40kg，问该地难产儿出生体重与一般新生儿体重有无差异？

本例已知总体均数 $\mu_0 = 3.30$kg，但总体标准差 σ 未知，$n = 35$ 为小样本，$\bar{X} = 3.42$kg，$S = 0.40$kg，出生体重一般可假设服从正态分布，故选用单样本 t 检验。

（1）建立检验假设，确定检验水准

$H_0: \mu = \mu_0$，该地难产儿与一般新生儿平均出生体重相同

$H_1: \mu \neq \mu_0$，该地难产儿与一般新生儿平均出生体重不同

$\alpha = 0.05$

（2）计算检验统计量

在 $\mu = \mu_0$ 成立的前提条件下，计算统计量为

$$t = \frac{\bar{X} - \mu_0}{S_{\bar{X}}} = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} = \frac{3.42 - 3.30}{0.40 / \sqrt{35}} = 1.77$$

（3）根据 P 值，作出推断结论

本例自由度 $\nu = n - 1 = 35 - 1 = 34$，查 t 界值表（附表 2），得 $t_{0.05/2,34} = 2.032$。因为 $t < t_{0.05/2,34}$，故 $P > 0.05$，表明差异无统计学意义，按 $\alpha = 0.05$ 水准不拒绝 H_0，即根据现有样本信息，尚不能认为该地难产儿与一般新生儿平均出生体重存在差异。

(李康,贺佳主编, 2024, p. 69) - Please click Zotero - Refresh in Word/LibreOffice to update all fields.

不拒绝 H_0 ≠ 证明没差异
只表示：**证据不够强，无法把“抽样波动”排除掉。**

二、配对样本均数 t 检验

配对 t 检验的本质不是“比两组”，
而是：
👉 比“同一对里的差值”是不是稳定地偏离 0。

1.什么时候必须用“配对 t 检验”

1️⃣ 数据结构的判断（第一步就决定成败）

你看到题目里有 **“一一对应”**， 就要警觉：

- 同一个人 **前 / 后**
- 同一标本 **两种处理**
- 按重要特征 **配成对子**（同性别、同年龄、同窝等）

📌 关键词：
同一个体 / 成对 / before–after / matched / paired

👉 **一旦是配对设计，就不能当作两独立样本。**

2.为什么配对 t 检验“更聪明”

普通两样本在干嘛？

- 把 **个体差异** 当噪声

配对设计在干嘛？

- **先在对子内部相减**
- 把个体差异直接消掉

📌 所以：

> **配对设计不是让样本更多，**
> **而是让噪声更小。**

3.配对 t 检验真正检验的是什么

关键转化（一定要会说）

配对 t 检验 ≡ **差值的单样本 t 检验**

1️⃣ 构造新变量：差值 d

对每一对数据：

$$d_i = X_{i1} - X_{i2}$$

你之后**只看**这组 d。

2️⃣ 检验假设（极其关键）

- 原假设（误伤防线）：

$$H_0 : \mu_d = 0$$

  👉 两种处理**平均没有差别**

- 备择假设：

$$H_1 : \mu_d \neq 0$$

📌 **直觉翻译**：

> 如果真没效果，那这些差值只是围着 0 抖。

4.统计量公式（只需要记一个）

$$t = \frac{\bar{d}}{S_d/\sqrt{n}}, v = n - 1$$

其中：

- d：每一对的差值
- $\bar{d}$：差值的样本均数
- S_d：差值的样本标准差
- n：对子数（⚠️ 不是原始观测数）

📌 **直觉一句话**:

> 平均差值，是“随机差值噪声”的几倍?

5.配对 t 检验【标准解题四步模板】

① 明确设计 & 选择方法

- 资料类型：配对设计计量资料
- 关注对象：对子内差值 d
- 方法：配对样本 t 检验（paired t-test）

---

② 建立假设 & 确定检验水准

$$H_0: \mu_d = 0$$
$$H_1: \mu_d \neq 0$$
$$\alpha = 0.05$$

---

③ 计算检验统计量

- 计算每对差值 d
- 求:
  - $\bar{d}$
  - S_d
- 计算:

$$t = \frac{\bar{d}}{S_d/\sqrt{n}}$$

- 自由度:

$$v = n - 1$$

---

④ 作出统计推断结论

- 若 $|t| \geq t_{\alpha/2,v}$ 或 $P < \alpha$
  👉 **拒绝 $\mathbf{H_0}$**
- 否则
  👉 **不拒绝 $\mathbf{H_0}$**

---

⑤ 规范结论表述

- **拒绝 $\mathbf{H_0}$**：

  在 $\alpha = 0.05$ 水准下，差异具有统计学意义，可以认为两种处理的效应存在差异。

- **不拒绝 $\mathbf{H_0}$**：

  在 $\alpha = 0.05$ 水准下，差异无统计学意义，尚不能认为两种处理的效应存在差异。

6、和单样本 / 两独立样本的“秒级区分法”

> **只问一句话就够了：这些数据之间有没有“一一绑定关系”？**

- 有 → **配对 t 检验**
- 没有 → 再考虑两独立样本 t / t′

【例 7-2】某项研究评估咖啡因对运动者心肌血流量的影响，先后测定了 12 名男性志愿者饮用咖啡前后运动状态下的心肌血流量[ml/(min·g)]，数据如表 7-1 所示，问饮用咖啡前后运动者的心肌血流量有无差异。

表 7-1 12 名运动者饮用咖啡前后的心肌血流量 单位：ml/(min·g)

编号	饮用前	饮用后	差值 d	编号	饮用前	饮用后	差值 d
1	4.8	4.8	0	7	5.1	4.1	1.0
2	5.1	4.9	0.2	8	4.9	3.2	1.7
3	6.4	4.5	1.9	9	4.7	3.0	1.7
4	5.7	5.4	0.3	10	3.5	3.2	0.3
5	5.6	4.7	0.9	11	5.2	5.3	−0.1
6	5.3	3.8	1.5	12	5.3	5.1	0.2

（1）建立检验假设，确定检验水准

H_0: $\mu_d=0$，饮用咖啡前后运动者的平均心肌血流量差值为零

H_1: $\mu_d\neq0$，饮用咖啡前后运动者的平均心肌血流量差值不为零

$\alpha=0.05$。

（2）计算检验统计量

先计算差值 d，如表 7-1 第四列所示。

本例 $n=12$，$\bar{d}=0.8$，$S_d=0.741$，计算差值的标准误：

$$S_{\bar{d}}=\frac{S_d}{\sqrt{n}}=\frac{0.741}{\sqrt{12}}=0.214$$

按公式（7-2）计算，得：

$$t=\frac{\bar{d}}{S_{\bar{d}}}=\frac{0.8}{0.214}=3.738$$

（3）根据 P 值作出推断结论

自由度 $\nu=n-1=12-1=11$，查 t 界值表（附表 2），得 $t_{0.05/2,11}=2.201$，本例 $t>2.201$，$P<0.05$，差别有统计学意义，拒绝 H_0，接受 H_1，可以认为饮用咖啡前后运动者的心肌血流量存在差异。

(李康,贺佳主编, 2024, p. 70) - Please click Zotero - Refresh in Word/LibreOffice to update all fields.

三、两独立样本均数 t 检验（two independent sample t-test)

两独立样本 t 检验问的是：

👉 两组“不同人”的平均差，是否大到不像是随机分组造成的？

1.什么时候用（第一眼判断法）

1️⃣ 数据结构是关键（比公式重要）

你看到题目里是：

- 两组 **不同个体**
- 完全随机分组
- **没有一一对应关系**

📌 关键词：

正常组 vs 实验组
男 vs 女
处理 A vs 处理 B
随机分成两组

👉 **这时才考虑“两独立样本 t 检验”**
⚠️ 一旦是前后、成对、匹配 → 立刻排除它。

2.为什么还能“转化成单样本 t 检验”（核心直觉）

关键视角（非常重要）

虽然你表面在比较两组：

- $\bar{X}_1$
- $\bar{X}_2$

但统计真正关心的是：

$$\bar{X}_1 - \bar{X}_2$$

而在原假设下：

$$H_0: \mu_1 = \mu_2 \Leftrightarrow \mu_1 - \mu_2 = 0$$

📌 所以本质上你在做的是：

“差值样本均数” vs “理论值 0” 的单样本 t 检验

3.t 值在干什么（一定要用这句话理解）

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

人话翻译：

两组均值差 ÷ 这个差在随机分组下本来会抖多大

- 分子：你看到的“组间差异”
- 分母：组内波动 + 样本量 → 形成的“噪声尺子”

👉 **差 / 噪声**，这是所有 t 检验的统一灵魂。

4.方差齐性为什么重要

1️⃣ 什么叫“方差齐”

- 两组的“个体波动程度”差不多
- 没有一组特别稳定、另一组特别乱

📌 只有在这种情况下，才敢用一个**合并方差**来估计噪声。

2️⃣ 合并方差 S_c^2在干嘛?

$$S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

直觉理解：

把两组的方差，
按各自样本量加权平均，
得到一个“共同噪声水平”。

3️⃣ 差值的标准误（噪声尺子）

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

📌 它体现了两点：

- 组内波动越大 → 噪声越大
- 样本量越大 → 噪声越小

5.两独立样本 t 检验【标准解题模板】

① 明确设计 & 选择方法

- 设计类型：完全随机设计
- 两组样本：相互独立
- 条件：正态分布 + 方差齐
- **方法**：两独立样本 t 检验

---

② 建立假设 & 检验水准

$$H_0: \mu_1 = \mu_2(\text{或}\mu_1 - \mu_2 = 0)$$
$$H_1: \mu_1 \neq \mu_2$$
$$\alpha = 0.05$$

📌 **直觉一句话**：默认“随机分组不会带来系统性差异”。

---

③ 计算检验统计量

- 合并方差：

$$S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- 差值标准误：

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

- t 值：

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1-\bar{X}_2}}$$

- 自由度：

$$v = n_1 + n_2 - 2$$

---

④ 作出统计推断结论

- 若 $|\, t \,| \geq t_{\alpha/2,v}$ 或 $P < \alpha$
  👉 **拒绝 H_0**
- 否则
  👉 **不拒绝 H_0**

---

⑤ 规范结论表述（考试友好）

- **拒绝 H_0**：

  在 $\alpha = 0.05$ 水准下，两组均数差异具有统计学意义。

- **不拒绝 H_0**：

  在 $\alpha = 0.05$ 水准下，两组均数差异无统计学意义，尚不能认为两总体均数不同。

6.什么时候**不能**用它（必考反例）

- 有前后对应 → ❌（应配对 t）
- 方差明显不齐 → ❌（应 t′ / Welch）
- 分布严重偏态、小样本 → ❌（需变换或非参数）

7、给你一句“统一三种 t 检验的终极理解”

单样本： 一组 vs 标准
配对样本： 同一对内先相减
两独立样本： 两组差值 vs 0

👉 **本质全是：差 ÷ 噪声**

【例 7-3】某项研究评估低氧环境(模拟高原环境)对运动者心肌血流量的影响,将 17 名男性志愿者随机分成两组,分别在正常含氧环境(正常组)和低氧环境(低氧组)中测定运动后的心肌血流量[ml/(min·g)],数据如表 7-2 所示,问两种环境中运动者的心肌血流量有无差异。

表 7-2　17 名运动者的心肌血流量　　单位:ml/(min·g)

正常组心肌血流量(X_1)	低氧组心肌血流量(X_2)	正常组心肌血流量(X_1)	低氧组心肌血流量(X_2)
3.5	6.4	2.3	4.9
3.1	5.7	2.3	4.7
3.1	5.6	2.2	3.5
2.7	5.3	2.2	
2.5	5.1		

(1)建立检验假设,确定检验水准

H_0: $\mu_1=\mu_2$,两种环境中运动者的心肌血流量的总体均数相同

H_1: $\mu_1\neq\mu_2$,两种环境中运动者的心肌血流量的总体均数不同

$\alpha=0.05$

(2)计算检验统计量

由原始数据计算得:

$$n_1=9,\quad \bar{X}_1=2.656,\quad s_1=0.475,\quad n_2=8,\quad \bar{X}_2=5.150,\quad s_1=0.852$$

代入公式(7-5)和公式(7-4),得:

$$S_c^2=\frac{(9-1)\times 0.475^2+(8-1)\times 0.852^2}{9+8-2}=0.459$$

$$S_{\bar{X}_1-\bar{X}_2}=\sqrt{0.459\times\left(\frac{1}{9}+\frac{1}{8}\right)}=0.329$$

(李康,贺佳主编, 2024, p. 71) - Please click Zotero - Refresh in Word/LibreOffice to update all fields.

按公式(7-3),算得

$$t=\frac{2.656-5.150}{0.329}=-7.581$$

(3)确定 P 值,作出推断结论

两独立样本 t 检验自由度为 $\nu=n_1+n_2-2=9+8-2=15$;查 t 界值表(附表 2),$t_{0.05/2,15}=2.131$。由于 $|t|>t_{0.05/2,15}$,$P<0.05$,按 $\alpha=0.05$ 水准双侧检验拒绝 H_0,接受 H_1,两组差异有统计学意义,可以认为两种环境中运动者的心肌血流量存在差异。

有些如抗体滴度的资料,宜用几何均数表示其平均水平。由于这些资料不服从正态分布(常服从对数正态分布),两样本的总体方差也可能不等,当对几何均数进行假设检验时,应先进行变量的对数变换,即将这些观察值 X 用 $\lg X$ 来代替,$\lg X$ 往往近似服从正态分布,相应的两总体方差也可能近似相等,故可用前述的 t 检验对 $\lg X$ 进行分析。

(李康,贺佳主编, 2024, p. 72) - Please click Zotero - Refresh in Word/LibreOffice to update all fields.

第二节 方差不齐时两样本均数的比较

一、方差齐性检验（F test）

> 方差齐性检验问的不是“均值”，
> 而是：**两组数据“乱得程度”是不是一个量级?**

1.为什么要先检验“方差齐不齐”（先讲动机）

你后面要用的 **两独立样本 t 检验** 有一个隐含前提：

> 👉 **两组的波动水平差不多**

否则：

- 用一个“合并方差”来当共同噪声
- 会 **低估或高估真实噪声**
- 直接影响 t 值 → 影响结论

📌 所以方差齐性检验是一个**前置安全检查**，不是主角。

2.方差齐性检验在比较什么（核心直觉）

你不是在比：

- 均值大不大

而是在比：

- **两组的“分散程度”**

如果：

- 两个总体方差真的相等
- 那从中抽样得到的两个样本方差
  👉 **应该只在 1 附近抖动**

3.F 统计量到底是什么（一定要这样记）

$$F = \frac{S_1^2(\text{较大})}{S_2^2(\text{较小})}$$

📌 三个关键点（考试必记）：

1️⃣ **大的放上面，小的放下面**
👉 所以 $F \geq 1$，不用纠结正负

2️⃣ 如果只是抽样误差
👉 F **不会离 1 太远**

3️⃣ **偏离 1 很多**
👉 才说明“波动程度不一样”

4.检验假设在说什么（用人话理解）

- 原假设：

$$H_0: \sigma_1^2 = \sigma_2^2$$

👉 两组“乱得一样”

- 备择假设：

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

👉 一组明显更乱 / 更稳

📌 **注意**：这是**双侧思想**
所以查表用 $\alpha/2$。

5.为什么常用 α=0.10（而不是 0.05）

这点教材常一笔带过，但你要有直觉。

> **方差齐性检验是“方法选择的筛查”，**
> **宁可保守一点，也别误用 t 检验。**

- 用 0.10：
  - 更容易发现“潜在不齐”
  - 避免在不齐的情况下硬用 pooled t

📌 **这是统计上的谨慎，不是随意。**

二、t' 检验｜方差不齐时的两独立样本均数比较

> **t′ 检验 = 两独立样本 t 检验的“修正版”。**
> 当两组波动程度不一样时：
> 👉 **别强行合并噪声，改用“各算各的噪声”。**

1.什么时候用 t'

触发条件（最常见）

- 两组 **独立样本**（不是配对）
- 大体可认为正态（或样本够大）
- **方差不齐**（通常先做 F 检验 / 或看数据明显不齐）

📌 直觉：

> 一组特别“乱”，一组特别“稳”，你再用同一个 pooled variance 当噪声尺子，会估错噪声。

2.t' 的核心变化是什么（直觉版）

1️⃣ t 的分子不变：还是均值差

$$\bar{X}_1 - \bar{X}_2$$

2️⃣ t 的分母换了：不再用合并方差

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

所以：

$$t' = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

📌 直觉一句话：

> **差 ÷（两组各自的均值波动，合起来的总噪声）**

3.为什么还要“校正”（自由度或临界值）

因为当方差不齐时：

- 你的噪声估计不再来自一个简单的 $v = n_1 + n_2 - 2$ 结构
- 所以 t 分布的“尾巴厚薄”需要调整

教材给两种路径：

路径 A：Satterthwaite（校正自由度）✅ 最常用

改自由度，临界值照查 t 表。

$$v' = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{(S_1^2/n_1)^2/(n_1-1) + (S_2^2/n_2)^2/(n_2-1)}$$

📌 直觉：

哪组方差更大、样本更小，越“拖后腿”，有效自由度就越低。

路径 B：Cochran & Cox（校正临界值）

自由度仍用各自的 v_1, v_2，把两边临界值加权平均成一个“校正临界值”。

先定义两组均值的方差（标准误平方）：

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

校正临界值：

$$t'_{\alpha/2} = \frac{S_{\bar{X}_1}^2 t_{\alpha/2,v_1} + S_{\bar{X}_2}^2 t_{\alpha/2,v_2}}{S_{\bar{X}_1}^2 + S_{\bar{X}_2}^2}$$

其中：

$$v_1 = n_1 - 1, v_2 = n_2 - 1$$

📌 直觉：

哪组均值更不稳定（标准误更大），就更影响临界值（权重大）。

【例 7-5】对例 7-4 资料进行 t' 检验，比较两组小白鼠增重均数是否不同。

（1）建立检验假设，确定检验水准

$H_0: \mu_1=\mu_2$，即两种饲料小白鼠增重总体均数相同

$H_1: \mu_1 \neq \mu_2$，即两种饲料小白鼠增重总体均数不相同

$\alpha=0.05$

（2）计算检验统计量

由表 7-3 数据算得：

$n_1=12$, $\bar{X}_1=45.750$, $S_1^2=17.659$

$n_2=13$, $\bar{X}_2=36.538$, $S_2^2=3.269$

由前面 F 检验得知两总体方差不同，因此应选用 t' 检验，即

$$t'=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}=\frac{45.750-36.538}{\sqrt{\dfrac{17.659}{12}+\dfrac{3.269}{13}}}=7.018$$

（3）确定 P 值，作出推断结论

按 Satterthwaite 法计算校正自由度，得

(李康,贺佳主编, 2024, p. 73) - Please click Zotero - Refresh in Word/LibreOffice to update all fields.

$$v'=\frac{(S_1^2/n_1+S_2^2/n_2)^2}{\dfrac{(S_1^2/n_1)^2}{n_1-1}+\dfrac{(S_2^2/n_2)^2}{n_2-1}}=\frac{(17.659/12+3.269/13)^2}{\dfrac{(17.659/12)^2}{12-1}+\dfrac{(3.269/13)^2}{13-1}}=14.687\approx 15$$

查 t 界值表（附表 2），$t_{0.05/2,15}=2.131$，$t'>t_{0.05/2,15}$，$P<0.05$。

按 Cochran & Cox 法计算校正界值，先查 t 界值表，得 $t_{0.05/2,11}=2.201$，$t_{0.05/2,12}=2.179$，再按公式（7-9），算得

$$t'_{0.05/2}=\frac{S_{\bar{X}_1}^2 t_{0.05/2,11}+S_{\bar{X}_2}^2 t_{0.05/2,12}}{S_{\bar{X}_1}^2+S_{\bar{X}_2}^2}=\frac{(17.659/12)\times 2.201+(3.269/13)\times 2.179}{(17.659/12)+(3.269/13)}=2.198$$

$t'>t'_{0.05/2}$，得 $P<0.05$。

两种检验方法所获得的界值虽略有差异，但结论是一致的。按 $\alpha=0.05$ 水准，拒绝 H_0，接受 H_1，两组差异有统计学意义，可认为两种饲料饲养后小白鼠增重的总体均数不同，即高蛋白组的体重增加量高于低蛋白组。

(李康,贺佳主编, 2024, p. 74) - Please click Zotero - Refresh in Word/LibreOffice to update all fields.

第三节 t 检验中的注意事项

> **一句总纲**：
> **t 检验不是魔法公式，**
> **它成立的前提，比算式本身更重要。**

一、结论成立的前提：样本 & 设计先过关

教材原意

> 样本要能代表总体，组间要均衡，设计要严谨。

人话直觉

> **统计不会拯救烂设计。**

你必须保证：

- 样本来自同质总体
- 抽样随机
- 分组在干预前完成（randomization）
- 组间差异主要来自处理，而不是背景因素

📌 否则你算出来的 t、P
只是**在错误前提上的精确数字。**

二、方法选对比算对更重要

t 检验在“吃什么条件”

- 近似正态分布（normality）
- 方差齐性（homogeneity of variance）
  ⚠️ **配对 t 不需要方差齐**

不满足怎么办？

- 方差不齐 → 用 t′（Welch / Satterthwaite）
- 分布偏态 → 变量变换（如 log / lg X）
- 样本量很大 → 可用 z 检验（CLT 兜底）

📌 **直觉一句话**：

> 别把 t 检验硬塞给不适合它的数据。

三、双侧 vs 单侧：这是“研究立场”，不是算术选择

关键原则

> **单双侧的选择，必须在看数据之前决定。**

双侧检验（two-sided）

- 不预设方向
- 更保守
- 默认首选

📌 大多数科研 & 考试场景 → **双侧**

单侧检验（one-sided）

- 只关心一个方向
- 常见于：
  - 非劣效性研究（non-inferiority）

    ◦ 安全性下限判断

⚠️ 在相同 α 下：

- 单侧 **更容易显著**
- 但 **更容易误伤**

📌 **直觉警告**：

> 事后改成单侧，本质是“换规则赢比赛”。

四、统计结论绝不能“说死”

正确态度

- 假设检验只给**概率保证**
- 不是绝对真理

报告时应：

- 给出 **确切 P 值**
- 说明 **单双侧检验**
- 小到极限可写：$P < 0.0001$

📌 **规范写作是学术诚信的一部分。**

五、P 值到底是什么意思（这是最容易被误解的点）

正确含义（必须记牢）

> **P 值 = 在 H_0 成立时，**
> **观察到当前或更极端差异的概率。**

❌ 不是：

- “差异是真的概率”
- “H_0 为假的概率”

P 值小 ≠ 差异大

从公式你已经看到:

- 样本量 n ↑
- 标准误 ↓
- t 值 ↑

📌 结果是:

> **只要样本足够大,**
> **再小的差也能显著。**

六、这一节真正想教你的“成熟统计观”

1️⃣ **先保证设计,再谈检验**
2️⃣ **方法选错,结果再显著也不可信**
3️⃣ **P 值回答“稳不稳”,不回答“大不大”**

[BIBLIOGRAPHY] Please click Zotero - Refresh in Word/LibreOffice to update all fields