

假设检验

第 11 讲 · 概率论与数理统计

复旦大学经济学院 ECON130001

上一讲我们做了什么

把「估计」这件事做完整了

- **三把尺子**：一致性、无偏性、有效性
- **点估计**：矩估计（矩对矩）、最大似然估计（哪个参数让数据最可能出现）
- **区间估计**：四个枢轴量 $\rightarrow \mu, \sigma^2, \mu_1 - \mu_2, \sigma_1^2/\sigma_2^2$ 的置信区间

★ 留下的问题

区间估计已经在悄悄做判断了：基金回报率的区间**整个在 0 以上**，我们说它赚钱；**跨过 0**，就说不出口。

可这个「说得出口」的界限，凭什么划在那里？**如果判断错了，错的概率是多少？**估计只回答「是多少」，**不回答「是不是」**。

为什么需要这一讲

场景 1920 年代，英国罗森斯特试验站的一次下午茶。一位藻类学家说：**奶茶是「先倒奶」还是「先倒茶」，一尝就能尝出来。** 在座的人笑了——两杯的化学成分完全一样。

冲突 有人提议：**那就倒一杯让她尝，尝对了就信她。** 听上去很公道。同在座的**费希尔**说：这个做法什么也证明不了。**就算完全靠猜，也有一半的概率蒙对。** 尝对一杯之后，你分不清「她真有本事」和「她运气好」——**这个实验从设计上就不可能给出答案。**

悬念 **那要尝几杯才够？「蒙对的概率足够小」——小到多少才算数？更要命的是：就算八杯全对，也没有「证明」她有这个本事，只能说「靠运气做到这一点太罕见了」。** **这两句话的差别在哪里？**

来源

R. A. Fisher, *The Design of Experiments* (1935) 第二章以此实验引出原假设概念。实验设计为 8 杯，其中 4 杯先奶后茶、4 杯先茶后奶，随机排列，受试者需指出是哪 4 杯。纯靠猜而全部答对的概率为 $1/\binom{8}{4} = 1/70 \approx 1.4\%$ 。据记载受试者全部答对。费希尔在同书中写道，原假设「never proved or established, but is possibly disproved, in the course of experimentation」。 en.wikipedia.org/wiki/Lady_tasting_tea | 访问日期 2026-08-02。

本讲学习目标

- ① 说清假设检验的**逻辑**：原假设、显著性水平、两类错误，以及「不拒绝」为什么不等于「接受」
- ② 掌握四种检验： z 、 t 、 χ^2 、 F ，会挑对统计量、查对临界值
- ③ 知道这套方法**会怎样被用坏**——以及为什么规则必须在看数据之前定好

与上一讲的分工

参数估计：对未知参数给出估计值或估计范围。

假设检验：事先对参数有一个预判，拿到样本后检验这个预判是否站得住。

「事先」两个字是要害——**先有假设，后看数据**。反过来做，整套理论就失效了（slide 24）。

假设检验的原理

三步

1. 把待检的假设写成 H_0 （原假设，也叫待检假设）
2. 在 H_0 成立的前提下，算出「出现手上这种样本」的概率
3. 如果这个概率很小，就认为 H_0 不合理，**拒绝** H_0 ； 否则**不能拒绝** H_0

「概率很小」小到多少

这是一个**主观选定**的标准，记为 α ，称为**显著性水平**（或检验水平），一般取 5% 或 1%。

★ 这个逻辑叫「小概率反证法」

它像数学里的反证法：先假定 H_0 成立，推出一个结论（这样的样本很罕见），再用观测到的事实去顶它。

但它不是严格的反证法：数学反证法推出的是矛盾，这里推出的只是不太可能。**不太可能的事也会发生**——所以这套程序注定会出错，只是出错的概率被我们事先控制住了。

两类错误

第一类错误：弃真

H_0 为真，却把它否定了。

发生概率就是显著性水平 α ——这是我们主动选定的。

第二类错误：取伪

H_0 为假，却没能否定它。

发生概率记为 β 。它不由我们直接选定，取决于真值离 H_0 有多远、样本有多大。

	H_0 实际为真	H_0 实际为假
拒绝 H_0	第一类错误（概率 α ）	判断正确
不拒绝 H_0	判断正确	第二类错误（概率 β ）

两者不能同时压小

样本容量固定时，把 α 调小， β 必然变大——门槛越高越不容易冤枉，也就越容易放过。标准做法是先固定 α 再降低 β ；而唯一能同时压小两者的，是加大样本量。

基本思想：为什么 H_0 受保护

新药审批

- **默认状态**：新药无效，不予上市
- **厂商的任务**：拿出足够强的临床证据，去推翻这个默认状态
- **结果一**：证据充分 → **批准**
- **结果二**：证据不足 → **不予批准**

假设检验

- **默认状态 H_0** ：无效应、无差异（新策略不赚钱、新算法没提升）
- **研究者的任务**：拿样本数据去挑战 H_0
- **结论一**：统计上显著 → **拒绝 H_0**
- **结论二**：证据不充分 → **不拒绝 H_0**

★ 最要紧的一条

「不予批准」**不等于**「证明了这个药无效」——它只意味着**手头的证据还不够**。很多药第一次试验没过，是因为样本量太小；加大试验后就过了。

同样，「不拒绝 H_0 」**绝不意味着**「证明了 H_0 正确」——所以标准说法是「**不拒绝原假设**」，而不是「**接受原假设**」。费希尔的原话是：原假设**永远不会被证明或确立**，只可能被推翻。

它到底在回答什么问题

这些问题每天都要用数据回答

- 新的广告投放策略，效果**真的**比旧的好吗？
- 这个量化交易模型，**真的**能赚钱吗（平均收益率大于 0）？
- 新版 App 的用户参与度，**真的**提高了吗？

句子里那两个字——「**真的**」——问的正是：**我们观测到的差异，是本质上的提升，还是随机波动造成的巧合？**

新版 App 的日均使用时长比旧版多了 3 分钟。这 3 分钟是真的吗？**只看这个数字永远回答不了**——因为哪怕两版完全一样，两组用户的均值也不可能恰好相等。

假设检验提供的，是一套**区分「随机巧合」与「本质不同」**的规则——并且明确告诉你，这套规则会以多大的概率出错。

单个正态总体均值： z 检验 (σ^2 已知)

检验 $H_0 : \mu = \mu_0$

样本来自 $N(\mu, \sigma^2)$, σ^2 已知。计算统计量

$$U = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

若 $|u| > u_{\alpha/2}$ 则拒绝 H_0 (临界值由 $P(|U| > u_{\alpha/2}) = \alpha$ 定)。

U 就是第 9 讲那个枢轴量, 只是把 μ 换成假设值 μ_0 。

例：砖的抗断强度

$\sigma^2 = 1.21$, 抽 6 块, 均值 31.13。 $\alpha = 5\%$ 下检验平均强度是否为 32.50。

$$u = \frac{31.13 - 32.50}{\sqrt{1.21/6}} = -3.05$$

$|u| = 3.05 > 1.96$, **拒绝** H_0 ——不能认为平均强度是 32.50。

例：钢索的断裂强度

$N(\mu, 40^2)$, 抽 9 根, 均值 780。 $\alpha = 0.05$ 下能否认为强度是 800?

$$u = \frac{780 - 800}{40 / \sqrt{9}} = -1.5$$

$|u| = 1.5 < 1.96$, **不能拒绝** H_0 。

注意措辞：结论是「**不能拒绝**」, 不是「已经证明断裂强度就是 800」。

两类错误的概率怎么算

仍是 σ^2 已知、检验 $H_0: \mu = \mu_0$ 的情形

第一类错误:

$$P(\text{拒绝} H_0 \mid H_0 \text{为真}) = P(|U| > u_{\alpha/2}) = \alpha$$

就是 α 本身——这正是我们选 α 的含义。

第二类错误 (真值是某个 $\mu \neq \mu_0$) :

$$\beta = P(|U| < u_{\alpha/2} \mid \mu) = \Phi_0(\lambda + u_{\alpha/2}) - \Phi_0(\lambda - u_{\alpha/2}), \quad \lambda = \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}$$

怎么读: λ 是真值与 μ_0 的距离, 按 $\sigma/\sqrt{n}$ 度量

- $\lambda \rightarrow 0$: $\beta \rightarrow 1 - \alpha$, **几乎必然发现不了**——差别太小, 本来就分不出
- $|\lambda|$ 很大: $\beta \rightarrow 0$, 差别明显, 几乎必然能发现
- n 变大 $\Rightarrow |\lambda|$ 变大 $\Rightarrow \beta$ 变小——**这就是加大样本量的全部作用**

$1 - \beta$ 称为**功效**。

单个正态总体均值： t 检验 (σ^2 未知)

σ^2 未知时用 S 顶替，统计量变成

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

若 $|t| > t_{\alpha/2}(n-1)$ 则拒绝 H_0 。与 z 检验只差两处： $\sigma \rightarrow S$, $u \rightarrow t$ 。

例：新生儿体重

抽 20 个，均值 3160、样本标准差 300；历史资料 3140。 $\alpha = 1\%$ 下有无显著差异 ($t_{0.005}(19) = 2.861$) ?

$$t = \frac{3160 - 3140}{300/\sqrt{20}} = 0.298$$

$|t| < 2.861$, 不能拒绝 H_0 ——没有显著差异。

例：元件寿命

16 只元件平均寿命 241.5、样本标准差 98.73，国标 280。 $\alpha = 5\%$ 下能否认为达标 ($t_{0.025}(15) = 2.131$) ?

$$t = \frac{241.5 - 280}{98.73/\sqrt{16}} = -1.560$$

$|t| < 2.131$, 不能拒绝 H_0 。

★ 右题值得多想一步

右题差了 38.5 (近 14%) 却「不能拒绝」——样本标准差高达 98.73，**差距淹没在噪声里。**「没有显著差异」不等于「没有差异」，它往往只是在说：数据还不足以分辨。

案例：这个交易策略真的能赚钱吗

场景

某日内交易策略：**不能稳定盈利的策略，日均收益率 μ 应为 0**。模拟 25 天， $\bar{x} = 0.12\%$ 、 $s = 0.5\%$ ，日收益率正态， $\alpha = 0.05$ 。

四步

1. **提出假设**： $H_0 : \mu \leq 0$ （不盈利）， $H_1 : \mu > 0$ （盈利）

2. **构造统计量**（ σ 未知，用 t ）：

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.12 - 0}{0.5/\sqrt{25}} = \frac{0.12}{0.1} = 1.2$$

3. **拒绝域**：单侧、 $df = 24$ ， $t_{0.05}(24) = 1.711$ ，即 $t > 1.711$

4. **决策**： $1.2 < 1.711$ ，**不能拒绝 H_0** ——没有充分证据说它赚钱

★ 为什么 H_0 是「不盈利」而不是「盈利」

因为 H_0 **是受保护的一方**：要说服人「这个策略能赚钱」，举证责任在声称赚钱的人身上——就像新药要证明自己有效。

把想证明的放进 H_1 ，把「什么也没发生」放进 H_0 ——这是列假设时唯一的规矩。

交互：看清两类错误

怎么用

左边设定 H_0 、真实世界的参数与检验参数，右边画出**两条抽样分布**—— H_0 假定的那条，和真实的那条。**两条曲线重叠的部分，就是错误发生的地方。**

当堂三个实验

1. 让真值 $= \mu_0$ ：拒绝的比例稳定在 α 附近——**第一类错误**
2. 让真值稍稍偏离 μ_0 ：大量该拒绝的没拒绝——**第二类错误**。再把真值拉远， β 迅速下降
3. 回到「真值稍稍偏离」，**只把样本量调大**：两条曲线都变窄、重叠变小， β 掉下去而 α 不变——**这就是加样本的作用**



交互演示：两条抽样分布的重叠部分，就是两类错误发生的地方

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/hypothesis_test.html

单个正态总体方差： χ^2 检验

检验 $H_0 : \sigma^2 = \sigma_0^2$ (μ 未知)

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

若 $\chi^2 > \chi_b^2$ 或 $\chi^2 < \chi_a^2$ 则拒绝 H_0 , 其中 $P(\chi^2 > \chi_b^2) = P(\chi^2 < \chi_a^2) = \alpha/2$, 自由度 $n-1$ 。

χ^2 不对称, 两个临界值要分别查表, 不是一对相反数。

例：铁水含碳量

改进工艺后抽 5 炉, 样本方差 0.228^2 。 $\alpha = 5\%$ 下能否认为方差仍为 0.108^2 ? ($\chi_{0.975}^2(4) = 0.484$, $\chi_{0.025}^2(4) = 11.1$)

$$\chi^2 = \frac{4 \times 0.228^2}{0.108^2} = 17.83$$

$17.83 > 11.1$, 拒绝 H_0 ——方差变了。

例：电池寿命的波动性

以往 $\sigma^2 = 5000$ 。新批次取 26 个, $s^2 = 9200$ 。 $\alpha = 2\%$ 下波动性有无显著变化? ($\chi_{0.99}^2(25) = 11.524$, $\chi_{0.01}^2(25) = 44.314$)

$$\chi^2 = \frac{25 \times 9200}{5000} = 46$$

$46 > 44.314$, 拒绝 H_0 ——波动性有显著变化。

单侧检验与一道综合题

$$H_0 : \sigma^2 \leq \sigma_0^2$$

统计量不变,

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

但只看**右尾**: $\chi^2 > \chi_\alpha^2(n-1)$ 就拒绝。

把 α 全押在一侧, 同样的显著性水平下更容易拒绝。

例: 包装机工作是否正常

每袋标准重量 500 g, 标准差不超过 10 g。抽 9 袋, 均值 499 g, 样本标准差 16.03 g。 $\alpha = 5\%$ ($t_{0.025}(8) = 2.306$, $\chi_{0.05}^2(8) = 15.5$)。

先检验均值 (σ 未知, 用 t) :

$$t = \frac{499 - 500}{16.03/\sqrt{9}} = -0.187$$

$|t| < 2.306$, **均值这一项正常**。

再检验方差 (单侧, $H_0 : \sigma^2 \leq 100$) :

$$\chi^2 = \frac{8 \times 16.03^2}{10^2} = 20.56$$

$20.56 > 15.5$, **拒绝——波动超标, 包装机工作不正常**。

一台机器要两项都合格。均值达标而方差超标, 意味着平均没错, 但每一袋都可能差很多——这比平均偏一点严重得多。

两个正态总体方差比: F 检验

检验 $H_0: \sigma_1^2 = \sigma_2^2$ (μ_1, μ_2 未知)

$$F = \frac{S_1^2}{S_2^2}$$

若 $F > F_b$ 或 $F < F_a$ 则拒绝, 其中 $P(F < F_a) = P(F > F_b) = \alpha/2$, 自由度 $(n_1 - 1, n_2 - 1)$ 。

H_0 为真时它退化成第 9 讲那个统计量, 应该在 1 附近; 单侧版本只看右尾。

例: 两种玉米的产量

甲、乙各 5 个地块, 样本方差 2653 与 11784,
 $\alpha = 5\%$ ($F_{0.025}(4, 4) = 9.60$ 、 $F_{0.975}(4, 4) = 0.10$)。

$$F = \frac{2653}{11784} = 0.23$$

$0.10 < 0.23 < 9.60$, **不能拒绝**——可以认为方差相等。

例: 两处煤矿的含灰率

甲 5 个方差 7.505, 乙 4 个方差 2.593, $\alpha = 10\%$
($F_{0.05}(4, 3) = 9.12$ 、 $F_{0.05}(3, 4) = 6.59$)。

$$F = \frac{7.505}{2.593} = 2.894. \text{ 左临界值要用倒数性质:}$$

$$F_{0.95}(4, 3) = \frac{1}{F_{0.05}(3, 4)} = \frac{1}{6.59} = 0.15$$

$0.15 < 2.894 < 9.12$, **不能拒绝**。

案例：两类资产的风险一样吗

场景

投资者在「新兴市场股票基金 A」与「发达国家债券基金 B」之间配置，想知道两者的风险（收益率方差）是否有显著差异。数据： $n_A = 16$ 、 $s_A^2 = 3.5$ ； $n_B = 21$ 、 $s_B^2 = 1.2$ 。 $\alpha = 0.05$ （ $F_{0.025}(15, 20) = 2.57$ ， $F_{0.975}(15, 20) = 0.363$ ）。

四步

1. $H_0 : \sigma_A^2 = \sigma_B^2$
2. $F = \frac{S_A^2}{S_B^2} \sim F(15, 20)$ ，观测值 $F = \frac{3.5}{1.2} \approx 2.92$
3. 拒绝域： $F > 2.57$ 或 $F < 0.363$
4. $2.92 > 2.57$ ，落入右侧拒绝域，**拒绝** H_0

怎么把结论说给客户听

有充分证据表明**两类资产的风险水平存在显著差异**，构建组合时不能把它们当作同风险资产对待。

两道单侧 F 检验

例：不同季节出生的新生儿体重

12 月出生抽 6 个，方差 505667；6 月出生抽 10 个，方差 93956。 $\alpha = 5\%$ 下检验**冬季的方差是否比夏季小** ($F_{0.05}(5, 9) = 3.48$)。

$H_0 : \sigma_1^2 \leq \sigma_2^2$ (冬季不比夏季大)：

$$F = \frac{505667}{93956} = 5.382 > 3.48$$

拒绝 H_0 ——冬季新生儿体重方差不比夏季小。

★ 单侧检验的假设要怎么摆

问题问的是「冬季是否比夏季小」，但 H_0 写的是**相反的那一边**。

规矩还是那条：**想证明的放 H_1** 。这里想证明「冬季方差更大」，所以 $H_1 : \sigma_1^2 > \sigma_2^2$ ， H_0 只能是 $\sigma_1^2 \leq \sigma_2^2$ 。**列反了拒绝域会跑到左尾**， $F = 5.382$ 落不进去，结论完全相反。

单侧的好处

本题做双侧的右临界值是 4.48，**单侧的门槛 3.48 更低**——方向说得越明确越容易显著。**所以方向必须在看数据之前定**：先看数据再挑方向，等于白拿一半的 α 。

两个正态总体均值差的检验

σ_1^2, σ_2^2 未知但相等, 检验 $H_0: \mu_1 = \mu_2$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

若 $|t| > t_{\alpha/2}(n_1 + n_2 - 2)$ 则拒绝 H_0 。

这就是第 9 讲那个两样本 t 统计量, 令 $\mu_1 - \mu_2 = 0$ 之后的样子。

★ 用它之前必须先做一件事

公式里的**合并方差** S_p^2 **假定了两个总体方差相等**。所以标准流程是**两步**:

1. 先用 F **检验**验证 $\sigma_1^2 = \sigma_2^2$ 是否站得住
2. 站得住, 才用上面这个 t 检验比较均值

顺序不能反。方差都不同的两组数据, 比较均值本来就没多大意义。

案例：App 新算法的 A/B 测试

场景

新版信息流推荐算法（B）是否比旧算法（A）更能增加用户停留时长？随机把 20 名用户分两组各 10 人，记录日均使用时长（分钟）。

A 组： $\bar{x}_A = 30.2$, $s_A^2 = 5.2$; B 组： $\bar{x}_B = 33.5$, $s_B^2 = 4.5$ 。 $\alpha = 0.05$ ($F_{0.05}(9, 9) = 3.179$, $t_{0.05}(18) = 1.734$)。

五步

1. **先查方差**: $F = \frac{5.2}{4.5} = 1.155$, 落在 $(1/3.179, 3.179)$ 内, 可以认为方差相等
2. **提假设**: $H_0 : \mu_A = \mu_B$, $H_1 : \mu_A < \mu_B$ (想证明 B 更好)
3. **合并方差**: $S_w^2 = \frac{9 \times 5.2 + 9 \times 4.5}{18} = 4.85$, $S_w \approx 2.20$
4. **统计量**: $t = \frac{30.2 - 33.5}{2.20 \sqrt{1/10 + 1/10}} \approx -3.35$
5. **决策**: 单侧拒绝域 $t < -1.734$, -3.35 落入其中, **拒绝** H_0

例题：单个总体的两种检验

例：考试成绩

$X \sim N(\mu, \sigma^2)$ ，抽 36 位考生，平均 66.5 分，标准差 15 分。 $\alpha = 0.05$ 下能否认为全体考生平均成绩为 70 分？

单个总体期望、 σ 未知，用 t ：

$$t = \frac{66.5 - 70}{15/\sqrt{36}} = \frac{-3.5}{2.5} = -1.4$$

临界值 $t_{0.025}(35) \approx 2.03$ ， $|t| < 2.03$ ，**不能拒绝** H_0 。

例：维尼纶纤度

正常条件下 $N(1.405, 0.048^2)$ 。某日抽 5 根，纤度为 1.32, 1.55, 1.36, 1.40, 1.44。 $\alpha = 0.1$ 下总体标准差是否正常？

$H_0 : \sigma^2 = 0.048^2$ 。算得 $\bar{x} = 1.414$ 、 $S^2 = 0.00778$ ：

$$\chi^2 = \frac{4 \times 0.00778}{0.048^2} = 13.5$$

临界值 $\chi_{0.05}^2(4) = 9.49$ 、 $\chi_{0.95}^2(4) = 0.711$ 。 $13.5 > 9.49$ ，**拒绝** H_0 ——标准差不正常。

两道题合起来是一个提醒：**题目问「均值」就用 t （或 z ），问「方差 / 标准差」就用 χ^2** 。右题给了 $\mu = 1.405$ 却用不上——检验方差时均值是不是已知都不影响统计量，因为 S^2 里已经用样本均值扣掉了它（代价是自由度少一个）。

例题：两个总体，先 F 后 t

例：两种橡胶配方的伸长率

A (9 个) : 540 533 520 545 531 541 529 534
525

B (10 个) : 565 577 580 575 556 542 560 532
570 561

$\alpha = 0.1$ 下**标准差**有无显著差异 ($F_{0.05}(9, 8) = 3.39$) ?

$S_1^2 = 63.86$ 、 $S_2^2 = 236.84$, $F = 0.2696$; 左临界值
 $F_{0.95}(8, 9) = 1/3.39 = 0.295$ 。 $0.2696 < 0.295$, **拒绝**。

例：两种提取工艺的得药率

旧: 75.5 77.3 76.2 78.1 74.3 72.4 77.4 78.4 76.7
76

新: 77.3 79.1 79.1 81 80.2 79.1 82.1 80 77.3
79.1

$\alpha = 0.01$ 下新旧工艺的期望有无显著差异?

① **查方差**: $s_1^2 = 3.325$ 、 $s_2^2 = 2.225$, $F = 1.49$ 落在
(1/6.54, 6.54) 内, **方差可视为相等**。

② **比均值**: $\bar{x} = 76.23$ 、 $\bar{y} = 79.43$ 、 $S_p = 1.666$,
 $T = \frac{76.23 - 79.43}{1.666\sqrt{0.2}} = -4.296$ 。 $|T| > t_{0.005}(18) =$
2.878, **拒绝**。

* 这些检验是从哪里来的：广义似然比

一般框架

总体密度 $f(x; \theta)$, $\theta \in \Theta$, 检验 $H_0 : \theta \in \Theta_0$ 。令

$$L(\hat{\Theta}) = \sup_{\theta \in \Theta} L, \quad L(\hat{\Theta}_0) = \sup_{\theta \in \Theta_0} L$$

称 $\lambda = \frac{L(\hat{\Theta})}{L(\hat{\Theta}_0)}$ 为**广义似然比**, 否定域为 $W_0 = \{\lambda > \lambda_0\}$, λ_0 由 $\sup_{\theta \in \Theta_0} P(W_0 | \theta) = \alpha$ 定出。

读法很直白：不受 H_0 约束时能达到的最大似然，比受约束时高出太多，就说明这个约束不合理。

双侧： $H_0 : \mu = \mu_0$ (σ^2 已知)

可算出 $\lambda = e^{\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2}$, 故

$$W_0 = \{|\bar{x} - \mu_0| > C\}$$

由 $P(|\bar{X} - \mu_0| > C | \mu_0) = \alpha$ 解得 $C = \frac{u_{\alpha/2}}{\sqrt{n}}\sigma$ ——

正是 slide 9 那个 z 检验。

单侧： $H_0 : \mu \leq \mu_0$

$\bar{x} \leq \mu_0$ 时 $\lambda \equiv 1$ (约束根本没起作用), 故只需看 $\bar{x} > \mu_0$, 此时 λ 与双侧同形, 但否定域变成

$$W_0 = \{\bar{x} - \mu_0 > C\}$$

由 $\sup_{\mu \leq \mu_0} P(\cdot) = P(\cdot | \mu_0) = \alpha$ 解得 $C = \frac{u_{2\alpha}}{\sqrt{n}}\sigma$ 。

交互：显著性是怎么被做出来的

这个工具里，原假设永远是真的

1. **「试很多次」**： $k = 20$ 个不相干的问题各检验一次，至少出一个假阳性的概率是 $1 - 0.95^{20} = 64\%$ ($k = 100$ 时 99.4%)。只报显著的那一次，读者以为是 5% 。
2. **「边看边停」**：每收一个新数据就检验一次，见到 $p < 0.05$ 就收工。 p 值在随机游走，迟早会掉下去。
3. **Bonferroni 校正**：门槛降到 α/k ，代价是 β 变大



$\alpha = 5\%$ 保证的是一次事先定好的检验犯第一类错误的概率是 5% ——这两个条件缺一不可。这也是第 4 讲那位「连胜十年」的基金经理的另一面。

本讲总结

① 回到那八杯茶

费希尔没有去争「她能不能尝出来」。他改而问了一个能回答的问题：**如果她完全靠猜，出现眼下这个结果的概率有多大？** 答案是 $1/70 \approx 1.4\%$ ——罕见到不像是猜的。

但即便八杯全对，费希尔也没有说「证明了她有这个本事」。他的原话是：原假设**永远不会被证明或确立，只可能被推翻**。

假设检验不是一台判定真假的机器。它做的是：把一个含糊的争论，换成一个有明确规则、并且**事先说好了会以多大概率出错**的判断程序。

本讲总结（续）

② 四种检验：挑对统计量

要检验什么	条件	统计量	分布
单个总体的 μ	σ^2 已知	$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$N(0, 1)$
	σ^2 未知	$\frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	$t(n - 1)$
单个总体的 σ^2	μ 未知	$\frac{(n - 1)S^2}{\sigma_0^2}$	$\chi^2(n - 1)$
两总体方差之比	μ_1, μ_2 未知	$\frac{S_1^2}{S_2^2}$	$F(n_1 - 1, n_2 - 1)$
两总体均值之差	方差未知但相等	$\frac{\bar{X} - \bar{Y}}{S_p\sqrt{1/n_1 + 1/n_2}}$	$t(n_1 + n_2 - 2)$

③ 易错提醒

- 比较两组均值前，先用 F 检验查方差；「不拒绝 H_0 」不是「接受 H_0 」
- χ^2 与 F 不对称，两端要分别查； F 的左临界值 = $1/F_\alpha(n, m)$ （取倒数、自由度换位）
- 没有显著差异 $\neq$ 没有差异；假设与方向必须事先定好

④ 下一讲

从分赌本到八杯茶，十二讲已全部讲完。**下一讲是复习课：**把这两半接成一张图，**看清两个方向在哪一点上对接。**