

第九章 假设检验

本章导读

参数估计回答“参数大概是多少”，假设检验回答“样本是否提供了足够证据，反对关于总体的某个说法”。其基本逻辑是反证式的：先暂时假定原假设成立，再判断当前样本对应的结果是否过于反常。

全章可归结为一条流程：

提出 H_0, H_1 $\longrightarrow$ 选检验统计量 $\longrightarrow$ 由 H_1 定方向 $\longrightarrow$ 由 α 定临界值 $\longrightarrow$ 作结论。

第八章与本章使用相同的正态、 t 、卡方和 F 枢轴量；区别只是前者解出参数区间，本章则判断观测统计量是否落入拒绝域。

9.1 假设检验的基本思想

9.1.1 原假设与备择假设

定义 9.1 (统计假设)

关于总体分布形式或未知参数的命题称为统计假设。检验时成对写为

H_0 ：原假设， H_1 ：备择假设。

H_0 是暂时保留、等待样本检验的命题； H_1 表示拒绝 H_0 后所支持的方向。

例如检验机器灌装均值是否仍为 500 克，可写

$H_0 : \mu = 500$, $H_1 : \mu \neq 500$.

若只担心灌装不足，则备择假设应改为 $H_1 : \mu < 500$ 。备择方向改变，拒绝域方向也随之改变。

9.1.2 小概率事件原理

假定 H_0 成立后, 选择一个分布已知的检验统计量. 若观测值落入在 H_0 下概率很小的区域, 就认为“ H_0 成立却出现如此极端数据”不够可信, 从而拒绝 H_0 .

注: [小概率不等于不可能] 小概率事件在一次试验中仍可能发生. 因此假设检验不能给出绝对正确的逻辑证明, 而是在预先控制错误风险的前提下作统计决策. 拒绝 H_0 可能犯错; 未拒绝 H_0 也不表示 H_0 已被证明为真.

定义 9.2 (拒绝域与接受域)

在 H_0 成立时, 检验统计量取值中会导致拒绝 H_0 的区域称为**拒绝域**; 其余区域常称为接受域或非拒绝域. 分隔二者的点称为临界值.

更准确的结论措辞是:

- 统计量落入拒绝域: 在显著性水平 α 下**拒绝** H_0 ;
- 统计量未落入拒绝域: 在显著性水平 α 下**没有充分理由拒绝** H_0 .

9.1.3 两类错误与显著性水平

实际情况	检验决定	结果
H_0 为真	不拒绝 H_0	正确
H_0 为真	拒绝 H_0	第一类错误
H_0 为假	拒绝 H_0	正确
H_0 为假	不拒绝 H_0	第二类错误

定义 9.3 (两类错误)

第一类错误是“弃真”, 其概率记为

$$\alpha = P(\text{拒绝}H_0 \mid H_0 \text{ 为真}).$$

第二类错误是“取伪”, 其概率通常记为

$$\beta = P(\text{不拒绝}H_0 \mid H_0 \text{ 为假}).$$

给定的 α 称为**显著性水平**. 本课程的显著性检验优先控制第一类错误概率不超过 α . 样本量

固定时, 不能通过简单移动临界值让 α 与所有备择参数下的 β 同时任意小; 增大样本量通常能改善两类风险.

9.1.4 标准检验步骤

1. 写出 H_0 与 H_1 , 明确是双侧、右侧还是左侧检验;
2. 在 H_0 成立条件下选择分布已知的检验统计量;
3. 根据显著性水平 α 和备择方向写拒绝域;
4. 代入样本观测值, 得到统计量的观测值;
5. 比较观测值与临界值, 并用完整句子作统计结论.

注: [备择假设决定拒绝域方向] 若 $H_1 : \theta \neq \theta_0$, 偏大或偏小都反对 H_0 , 所以拒绝域在两侧, 每侧概率为 $\alpha/2$; 若 $H_1 : \theta > \theta_0$, 只有统计量过大才反对 H_0 , 拒绝域在右侧; 若 $H_1 : \theta < \theta_0$, 拒绝域在左侧. 方向来自备择假设, 不来自样本观测值.

小结

假设检验先假定 H_0 成立, 再用小概率事件原理判断样本是否过于极端. H_1 决定拒绝域方向, α 控制第一类错误并决定临界值; 未落入拒绝域只能说证据不足, 不能证明 H_0 为真.

9.2 双侧、单侧检验与 p 值

9.2.1 三种拒绝域的统一形式

设在 H_0 下, 检验统计量 T 的分布关于 0 对称, 上 γ 分位点记为 t_γ . 则

备择假设	检验类型	拒绝域
$H_1 : \theta \neq \theta_0$	双侧	$ T > t_{\alpha/2}$
$H_1 : \theta > \theta_0$	右侧	$T > t_\alpha$
$H_1 : \theta < \theta_0$	左侧	$T < -t_\alpha$

对卡方、 F 等不对称且只取正值的分布, 左侧临界值应写成上 $1 - \alpha$ 分位点. 例如左侧 F 检验拒绝域为 $F < F_{1-\alpha}$.

例题 9.4 (备择方向不能由数据临时决定)

研究新工艺是否**提高**平均强度, 应事先写 $H_1 : \mu > \mu_0$, 采用右侧检验. 即使样本均值恰好比 μ_0 小, 也不能看完数据后改成左侧检验; 这种事后改方向会破坏预先设定的第一类错误概率.

9.2.2 p 值

定义 9.5 (p 值)

在 H_0 成立的前提下, 得到“至少与当前观测结果同样极端”的统计量取值的概率, 称为该检验的 p 值. 等价地, p 值是刚好能够拒绝 H_0 的最小显著性水平.

决策规则为

$$p \leq \alpha \Rightarrow \text{拒绝 } H_0, \quad p > \alpha \Rightarrow \text{不拒绝 } H_0.$$

对标准正态检验, 若观测值为 z_0 , 则

$$H_1 : \mu \neq \mu_0 \Rightarrow p = 2P(Z \geq |z_0|),$$

$$H_1 : \mu > \mu_0 \Rightarrow p = P(Z \geq z_0),$$

$$H_1 : \mu < \mu_0 \Rightarrow p = P(Z \leq z_0).$$

注: p 值不是“ H_0 为真的概率”, 也不是“本次结论犯错的概率”. 它是在 H_0 已被假定成立时, 当前数据及更极端数据出现的尾概率. p 值越小, 数据与 H_0 越不相容, 但效应大小还应结合实际背景判断.

9.2.3 与置信区间的联系

对使用同一枢轴量的双侧检验, 显著性水平为 α 的检验与置信水平为 $1 - \alpha$ 的置信区间具有对应关系:

$$\theta_0 \notin (1 - \alpha) \text{ 置信区间} \iff \text{拒绝 } H_0 : \theta = \theta_0.$$

若 θ_0 落在区间内, 则不能拒绝 H_0 . 这种等价关系主要针对双侧检验, 且必须使用相同模型、枢轴量和显著性水平.

小结

双侧检验把 α 平分到两端, 右侧和左侧检验把全部 α 放在备择假设指向的一端. p 值用观测结果对应的尾概率度量反对 H_0 的证据强弱; 双侧检验还可用对应置信区间作等价判断.

9.3 一个正态总体均值的检验

设 $X_1, \dots, X_n$ 来自 $N(\mu, \sigma^2)$, 要检验

$$H_0 : \mu = \mu_0.$$

9.3.1 方差已知: Z 检验

在 H_0 下,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

定理 9.6 (正态均值的 Z 检验)

显著性水平为 α 时, 拒绝域为

H_1	拒绝域
$\mu \neq \mu_0$	$ Z > z_{\alpha/2}$
$\mu > \mu_0$	$Z > z_\alpha$
$\mu < \mu_0$	$Z < -z_\alpha$

例题 9.7 (完整的“统计量—临界值—结论”)

已知某正态总体标准差 $\sigma = 6$, 现取 $n = 36$, 得到 $\bar{x} = 102.5$. 在 $\alpha = 0.05$ 下检验

$$H_0 : \mu = 100, \quad H_1 : \mu > 100.$$

这是右侧 Z 检验, 统计量观测值为

$$z_0 = \frac{102.5 - 100}{6/\sqrt{36}} = 2.5.$$

右侧临界值 $z_{0.05} = 1.645$. 因为 $2.5 > 1.645$, 观测值落入拒绝域, 所以在 0.05 显著性水平下拒绝 H_0 , 认为样本提供了总体均值大于 100 的显著证据.

结果检查: 备择为“大于”, 拒绝域必须在右侧; 结论不能写成“证明 $\mu > 100$ ”.

9.3.2 方差未知: t 检验

若 σ^2 未知, 在 H_0 下使用

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

定理 9.8 (正态均值的 t 检验)

显著性水平为 α 时, 拒绝域为

H_1	拒绝域
$\mu \neq \mu_0$	$ T > t_{\alpha/2}(n-1)$
$\mu > \mu_0$	$T > t_{\alpha}(n-1)$
$\mu < \mu_0$	$T < -t_{\alpha}(n-1)$

注: [Z 与 t 怎样选] 选择标准只看总体方差是否已知, 不看样本量“大不大”. 本课程的精确结论建立在正态总体上: 方差已知用 σ 和 Z 检验; 方差未知用 S 和自由度为 $n-1$ 的 t 检验.

小结

一个正态总体均值检验以 $\bar{X} - \mu_0$ 为分子; 方差已知标准化为 Z, 方差未知学生化为 t. 备择假设为不等、大于、小于时, 分别使用双侧、右侧、左侧拒绝域.

9.4 一个正态总体方差的检验

要检验

$$H_0 : \sigma^2 = \sigma_0^2.$$

当 μ 已知时, 使用

$$Q = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \sim \chi^2(n).$$

当 μ 未知时, 使用

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

下表用 k 表示相应自由度.

定理 9.9 (正态方差的卡方检验)

显著性水平为 α 时, 拒绝域为

H_1	拒绝域
$\sigma^2 \neq \sigma_0^2$	$Q < \chi_{1-\alpha/2}^2(k)$ 或 $Q > \chi_{\alpha/2}^2(k)$
$\sigma^2 > \sigma_0^2$	$Q > \chi_{\alpha}^2(k)$
$\sigma^2 < \sigma_0^2$	$Q < \chi_{1-\alpha}^2(k)$

其中 μ 已知时 $k = n$, μ 未知时 $k = n - 1$.

例题 9.10 (不对称分布的双侧拒绝域)

正态总体均值未知, $n = 20$, 要在显著性水平 0.05 下检验 $H_0 : \sigma^2 = 9$ 对 $H_1 : \sigma^2 \neq 9$.

若样本方差为 s^2 , 统计量为

$$q_0 = \frac{19s^2}{9}.$$

拒绝域是

$$q_0 < \chi_{0.975}^2(19) \quad \text{或} \quad q_0 > \chi_{0.025}^2(19).$$

易错提醒: 卡方分布不对称, 不能写成 $|Q| > c$; 左侧小临界值的上分位点下标是 0.975.

小结

正态方差检验使用卡方统计量. 均值已知时自由度为 n , 未知时为 $n - 1$; 双侧拒绝域位于两个不对称尾部, 右侧和左侧检验分别使用上 α 与上 $1 - \alpha$ 分位点.

9.5 两个正态总体均值差的检验

设两组独立样本分别来自 $N(\mu_1, \sigma_1^2)$ 与 $N(\mu_2, \sigma_2^2)$, 检验

$$H_0 : \mu_1 - \mu_2 = \delta_0.$$

9.5.1 两个方差已知

在 H_0 下,

$$Z = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1).$$

双侧、右侧、左侧拒绝域分别为

$$|Z| > z_{\alpha/2}, \quad Z > z_{\alpha}, \quad Z < -z_{\alpha}.$$

9.5.2 两个方差未知但相等

若 $\sigma_1^2 = \sigma_2^2$ 未知, 定义

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

则在 H_0 下

$$T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S_w \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2).$$

双侧、右侧、左侧拒绝域分别为

$$|T| > t_{\alpha/2}(n_1 + n_2 - 2),$$

$$T > t_{\alpha}(n_1 + n_2 - 2), \quad T < -t_{\alpha}(n_1 + n_2 - 2).$$

例题 9.11 (检验两个工艺的平均值是否不同)

两组独立正态样本方差未知但可认为相等, 要检验

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2.$$

应取 $\delta_0 = 0$, 先由两组样本方差计算 S_w^2 , 再计算

$$t_0 = \frac{\bar{x} - \bar{y}}{s_w \sqrt{1/n_1 + 1/n_2}}.$$

若 $|t_0| > t_{\alpha/2}(n_1 + n_2 - 2)$, 则拒绝“两个均值相等”; 否则只能说尚无充分证据认为两个均值不同.

小结

两个均值差检验与其置信区间使用同一枢轴量. 两个方差已知用 Z ; 未知但相等用合并方差和自由度 $n_1 + n_2 - 2$ 的 t . 两组样本独立和等方差条件都不能漏掉.

9.6 两个正态总体方差比的检验

检验一般形式

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = r_0, \quad r_0 > 0.$$

在 H_0 下,

$$F = \frac{S_1^2/S_2^2}{r_0} \sim F(n_1 - 1, n_2 - 1).$$

定理 9.12 (两个正态总体方差比的 F 检验)

显著性水平为 α 时, 拒绝域为

H_1	拒绝域
$\sigma_1^2/\sigma_2^2 \neq r_0$	$F < F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$ 或 $F > F_{\alpha/2}(n_1 - 1, n_2 - 1)$
$\sigma_1^2/\sigma_2^2 > r_0$	$F > F_{\alpha}(n_1 - 1, n_2 - 1)$
$\sigma_1^2/\sigma_2^2 < r_0$	$F < F_{1-\alpha}(n_1 - 1, n_2 - 1)$

最常见的原假设是 $r_0 = 1$, 即 $H_0 : \sigma_1^2 = \sigma_2^2$, 此时检验统计量就是 S_1^2/S_2^2 .

注: [自由度顺序与比值顺序绑定] 分子 S_1^2 对应第一个自由度 $n_1 - 1$, 分母 S_2^2 对应第二个自由度 $n_2 - 1$. 若把统计量改写成倒数, 两个自由度和备择方向都要相应交换.

小结

方差比检验使用 F 分布. 原假设中的比值 r_0 要除入统计量; 双侧拒绝域在两个不对称尾部, 单侧方向仍由备择假设决定. 书写临界值时必须保持样本比值与自由度顺序一致.

9.7 统计结论与数学证明的边界

9.7.1 未拒绝不等于证明为真

一次样本未落入拒绝域, 可能因为 H_0 确实成立, 也可能因为样本量不足、数据波动较大而未能识别差异. 因此只能说“没有充分理由拒绝 H_0 ”, 不能说“接受并证明 H_0 为真”.

同理, 拒绝 H_0 表示数据在 H_0 下过于极端, 并承担至多 α 的第一类错误风险; 它不是演绎逻辑中的绝对否定.

9.7.2 反例与普遍命题

假设检验的证据逻辑不能替代数学命题的证明规则. 对全称命题“所有对象都具有性质 P ”:

- 找到一个不具有 P 的对象, 就能用反例否定全称命题;
- 检查若干对象都具有 P , 只能增加直观信心, 不能证明所有对象都具有 P ;
- 要证明普遍命题, 仍需从定义和已知条件作一般性推导.

例如, 要否定“任意两个不相关随机变量都独立”, 给出一个不相关但不独立的反例即可; 列举多个独立且不相关的例子, 却不能证明该命题成立.

小结

统计检验给出带错误风险的证据判断, 不是对参数命题的绝对证明. 数学上否定全称命题可用一个反例, 证明全称命题则必须覆盖一般对象; 这两种逻辑边界都应明确.

本章总结

一、统一决策流程

$H_0, H_1 \longrightarrow$ 在 H_0 下确定统计量分布 $\longrightarrow$ 按 H_1 定尾部 $\longrightarrow$ 按 α 定临界值 $\longrightarrow$ 比较并作结论.

二、正态总体检验选择表

待检参数	条件	检验统计量	原假设下分布
单个 μ	σ^2 已知	$(\bar{X} - \mu_0)/(\sigma/\sqrt{n})$	$N(0, 1)$
单个 μ	σ^2 未知	$(\bar{X} - \mu_0)/(S/\sqrt{n})$	$t(n-1)$
单个 σ^2	μ 已知	$\sum (X_i - \mu)^2/\sigma_0^2$	$\chi^2(n)$
单个 σ^2	μ 未知	$(n-1)S^2/\sigma_0^2$	$\chi^2(n-1)$
$\mu_1 - \mu_2$	两方差已知	标准化均值差	$N(0, 1)$
$\mu_1 - \mu_2$	两方差未知但相等	合并方差学生化均值差	$t(n_1 + n_2 - 2)$
σ_1^2/σ_2^2	两组独立正态样本	$(S_1^2/S_2^2)/r_0$	$F(n_1-1, n_2-1)$

三、方向与临界值速查

备择方向	对称分布	卡方或 F 分布	含义
$\neq$	$ T > c_{\alpha/2}$	小于上 $1 - \alpha/2$ 或大于上 $\alpha/2$ 分位点	双侧
$>$	$T > c_{\alpha}$	大于上 α 分位点	右侧
$<$	$T < -c_{\alpha}$	小于上 $1 - \alpha$ 分位点	左侧

四、重点辨析与易错点

- 备择假设决定拒绝域方向, 不能看完样本后再改成更容易拒绝的方向.
- 显著性水平是第一类错误概率的控制界, 不是 H_0 为假的概率.
- p 值是在 H_0 下的尾概率, 不是结论出错概率; $p \leq \alpha$ 才拒绝 H_0 .
- 未拒绝 H_0 不等于接受或证明 H_0 ; 结论必须保留“证据是否充分”的措辞.
- 卡方与 F 的双侧临界值不对称; 使用上分位点时, 下标顺序不能照搬对称分布.
- Z 、 t 、卡方、 F 的选择条件与上一章置信区间完全对应.

五、证明掌握清单

- **A 级, 应能独立完成:** 根据 H_0 选择检验统计量; 根据 H_1 推出双侧、左侧或右侧拒绝域; 说明 α 如何决定临界值; 完成“统计量—临界值—结论”的完整论证; 解释为何不能由未拒绝证明 H_0 ; 用反例否定全称命题.
- **B 级, 应理解思路:** 假设检验与置信区间使用相同枢轴量的联系; 两类错误与显著性水平; 单侧检验方向为何由备择假设决定.
- **C 级, 知道结论即可:** 各种拒绝域公式的重复代数推导; 检验最优性、功效函数与 Neyman–Pearson 理论. 这些内容本阶段不要求复现.

六、自测题

练习 9.13

分别说明第一类错误和第二类错误, 并解释为什么显著性水平 α 主要控制第一类错误.

练习 9.14

正态总体方差已知, 写出检验 $H_0: \mu = \mu_0$ 对三种备择 $\mu \neq \mu_0$ 、 $\mu > \mu_0$ 、 $\mu < \mu_0$ 的拒绝域.

练习 9.15

某检验得到 p 值 0.032. 分别在 $\alpha = 0.05$ 和 $\alpha = 0.01$ 下作结论, 并说明 p 值不能解释为什么.

练习 9.16

正态总体均值未知, 说明检验 $H_0: \sigma^2 = \sigma_0^2$ 时统计量、自由度和三类拒绝域.

练习 9.17

两组独立正态样本方差未知但相等, 写出检验 $H_0: \mu_1 - \mu_2 = 0$ 的统计量, 并说明双侧临界值与自由度.

练习 9.18

构造一个反例, 否定“任意两个不相关随机变量都相互独立”, 并说明为什么若干正例不能证明该全称命题.

完成自测后, 应能用一句话概括全章: 在预先控制第一类错误的前提下, 判断样本是否与原假设过于不相容, 并作出拒绝或不拒绝的证据结论.