概率论与数理统计

第八章: 假设检验 (Testing of Hypothesis)

李立颖 (lily@sustech.edu.cn)

南方科技大学数学系

2023 秋季

- 参数估计问题：参数估计是通过样本的观察对总体分布中的参数作出估计。
- **假设检验问题:** 就是要根据样本的信息来判断总体分布是否有理由说它是来自均值为 μ_0 的正态总体，是否能说这两个总体的均值相同或方差相同？
  本章主要考虑以样本均值和样本方差为工具的一些较为重要的假设检验。

例

甲公司从乙工厂在订购了一批元件，合同规定产品次品率不得超过 5%，否则甲方有权拒收. 令从一大批元件中抽检 100 件，发现次品 4 件，问甲方是否应该接收这批元件?

分析

求得元件次品率 p 的极大似然估计值应该为 $\hat{p} = \bar{X} = 0.04 < 5\%$.

甲方顾虑　因 $\bar{X}$ 是随机变量，其值带有波动性，有可能 $\bar{X} < 5\%$ 但次品率实际上超过了 5%.

乙方想法
1. $\hat{p} = \bar{X}$ 是 p 的极大似然估计，且 $\mathbb{E}\hat{p} = p$ (无偏)，
   $$\mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n} \to 0,\ n \to \infty.$$
2. $\bar{x} = 4\% < 5\%$

故工厂无法接受 “产品不能出厂” 的结论.

体重指数 BMI

BMI 是国际上常用的衡量人体胖瘦程度以及是否健康的一个标准. 专家指出, 健康成年人的 BMI 取值应该在 18.55 – 24.99 之间.

某种减肥药广告宣称, 连续使用该种减肥花一个星期便可达到减肥的效果.
为了检验其说法是否可靠, 随机抽取 9 位试验者 (BMI 指数超过 25, 年龄在 20 – 25 岁之间女生), 先让每位女生记录没有服用减肥药之前的体重, 然后让每位女生服用该减肥药. 在服药期间, 要求每位女生保持正常的饮食习惯, 连续服用一周后, 再次记录各自的体重. 测得服药前后的体重减小的值为 (kg)

$$1.5,\ 0.6,\ -0.3,\ 1.1,\ -0.8,\ 0,\ 2.2,\ -1.0,\ 1.4$$

问题

根据目前的样本资料, 能否认为该减肥药广告中的宣称是可靠的?

第一步：建立假设

建立两个完全对立的假设：

原假设/零假设：H_0；　备择假设/对立假设：H_1.

原假设与备择假设是不对称的！决定谁是原假设，依赖于立场、惯例、方便性.

保护原假设

如果错误地拒绝假设 A 比错误地拒绝假设 B 带来更严重的后果 — A 选作原假设！例：

假设 A：新药有某种毒副作用，　假设 B：新药无某种毒副作用 — A 选作原假设 H_0.

这是因为，“有毒副作用”错误地当成“无毒副作用”比“无毒副作用”错误地当成“有毒副作用”带来的后果更严重.

原假设为维持现状

为解释某些现象或效果的存在性，原假设常取为“无效果”、“无改进”、“无差异”等，拒绝原假设表示有较强的理由支持备择假设. 例：

H_0：药物没有减肥效果，　H_1：药物有减肥效果.

第一步：建立假设

原假设取简单假设

只有一个参数 (或分布) 的假设称为**简单假设**. 如果只有一个假设是简单假设, 将其取为原假设.

参数假设的形式

设 θ 是反映总体指标某方面特征的量, 是我们感兴趣的参数. 一般参数 θ 的假设有三种情形:

左边检验 $H_0:\theta=\theta_0,\ H_1:\theta<\theta_0.$

右边检验 $H_0:\theta=\theta_0,\ H_1:\theta>\theta_0.$

双边检验 $H_0:\theta=\theta_0,\ H_1:\theta\neq\theta_0.$

注

在假设检验中,

$$H_0:\theta\geq\theta_0,\ H_1:\theta<\theta_0, \quad \text{与} \quad H_0:\theta=\theta_0,\ H_1:\theta<\theta_0$$

的检验法则与检验效果是一样的. 另一个方向同理.

例

某工厂生产的一种产品平均寿命为 2000 小时，生产车间采用一种新工艺后，认为产品寿命有了显著提高．现从采用新工艺生产的产品中随机抽检了 25 件产品，试验后算得寿命平均值为 2085 小时．问能否认为新工艺能显著提高产品质量?

分析

设产品寿命 $X \sim \mathcal{N}(\mu, \sigma^2)$. μ 的点估计为 $\hat{\mu} = \bar{x} = 2085 > 2000$ 小时，故认为新工艺能提高质量的看法是有依据的．

问题

点估计 $\hat{\mu} = \bar{X}$ 是随机变量，取值具有波动性，故尽管 $\bar{x} = 2085 > 2000$，但产品质量可能并无显著提高．

- 提出假设: $H_0 : \mu = 2000$, $H_1 : \mu > 2000$.
- 作为判断:
  - 拒绝 H_0 即认为产品寿命有显著提高．
  - 接受 H_0 即认为产品寿命没有显著提高．

第二步：给出检验统计量，并确立拒绝域的形式

如何检验假设

根据收集的资料，针对假设，给出检验方法，然后对假设进行判断．有两种方法：**临界值法**，p **值法**．

设服用减肥药前后体重差值 $X \sim \mathcal{N}(\mu, \sigma^2)$，并假定方差为 $\sigma^2 = 0.36$.

$$\textbf{检验假设:}\ H_0: \mu = 0,\ H_1\colon\ \mu > 0.$$

注意到 $\bar{X}$ 是 μ 的无偏估计，$\bar{X}$ 的取值大小反映了 μ 的取值大小，当原假设成立时，$\bar{X}$ 的取值应偏小．因此我们提出以下检验方法：

- 当 $\bar{X} \geq C$ 时，拒绝原假设 H_0，
- 当 $\bar{X} < C$ 时，接受原假设 H_0.

C 为待定常数．

如果统计量 $T = T(X_1, \ldots, X_n)$ 的取值大小和原假设 H_0 的是否成立有密切关系，可将其称为对应假设问题的**检验统计量**，而对应于拒绝于原假设 H_0 时，样本值的范围称为**拒绝域**，记为 W；其补集 $\bar{W}$ 称为**接受域**.
上例中的检验统计量为 $\bar{X}$，拒绝域为 $W = \{(X_1, \ldots, X_n) : \bar{X} \geq C\}$.

第一类与第二类错误

由于样本的随机性，任一检验规则在应用时，都有可能发生错误的判断

第 I 类错误 拒绝真实的原假设 (弃真)

第 II 类错误 接受错误的原假设 (取伪).

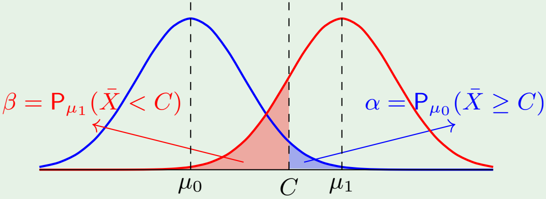
我们考察两类错误发生的概率:

$$\alpha = \mathsf{P}(\text{ Type I error}) = \mathsf{P}(\text{拒绝 } H_0 \mid H_0 \text{ 是真实的}),$$
$$\beta = \mathsf{P}(\text{ Type II error}) = \mathsf{P}(\text{ 接受 } H_0 \mid H_0 \text{ 是错误的}).$$

例

总体 $X \sim \mathcal{N}(\mu, 1)$，则 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{1}{n})$. 令 $H_0: \mu = \mu_0$，$H_1: \mu = \mu_1 \ (> \mu_0)$，

拒绝域：$\bar{X} \geq C$.



两类错误相互制约！

Neyman–Pearson 原则

首先控制犯第 I 类错误的概率不超过某个常数 $\alpha \in (0,1)$，再寻找检验，使得犯第 II 类错误的概率尽可能小. α 称为**显著水平**.
常取 $\alpha = 0.01, 0.05, 0.1$ 等.

第三步：根据显著水平和统计量的分布确定临界值 — 临界值法

在减肥药的例子中，取显著水平 $\alpha = 0.05$，则

$$H_0 : \mu = 0 \text{ 成立} \quad \Rightarrow \quad \frac{\bar{X}}{0.6/\sqrt{9}} \sim \mathcal{N}(0,1),$$

于是第 I 类错误的概率可如下计算：

$$\mathsf{P}(\bar{X} \geq C \mid \mu = 0) = \mathsf{P}\left(\frac{\bar{X}}{\sigma/\sqrt{n}} \geq \frac{C}{\sigma/\sqrt{n}} \mid \mu = 0\right) = 1 - \Phi\left(\frac{C}{\sigma/\sqrt{n}}\right) \leq \alpha = 0.05$$

$$\Rightarrow \frac{C}{0.6/\sqrt{9}} \geq u_{0.05} = 1.645 \quad \Rightarrow \quad C \geq 0.329.$$

根据 Neyman–Pearson 原则，为使犯第 II 类错误的概率尽可能小，应取 $C = 0.329$. 因此，我们选取拒绝域为

$$W = \{(X_1, \ldots, X_n) : \bar{X} \geq 0.329\}.$$

第四步：根据样本得出结论

根据实际样本资料，得 $\bar{x} = 0.522 > 0.329$.
当原假设 H_0 成立时，样本落在拒绝域的概率不超过 0.05，是**小概率事件**.
根据**实际推断原理，有充分的理由拒绝原假设**，因此我们认为厂家的宣传是可靠的.
同理，若 $\alpha = 0.01$，拒绝域 $W = \{\bar{X} \geq 0.465\}$，此时我们拒绝原假设.

练习

例 1

假设检验的统计思想是小概率事件在一次试验中可能认为基本上是不会发生的，该原理称为：

实际推断原理.

例二

假设检验中，第 I 类和第 II 类错误分别指：

第 I 类：弃真；第 II 类：取伪.

练习

例三

假设检验中，显著性水平 α 表示

(A) H_0 为假，但接受 H_0 的假设的概率；

(B) H_0 为真，但拒绝 H_0 的假设的概率；

(C) H_0 为假，但拒绝 H_0 的概率.

(D) 可信度.

B.

例四

假设检验时，若增大样本容量，则犯两类错误的概率

(A) 都增大；

(B) 都减少；

(C) 都不变；

(D) 一个增大，一个减少.

假设检验

- 单正态总体在 σ^2 已知的情况下，关于 μ 的检验问题.
  - 双边 u 检验法
  - 单边 u 检验法
- 单正态总体在 σ^2 未知的情况下，关于 μ 的检验问题
  - 双边 t 检验法
  - 单边 t 检验法
- 单正态总体在 μ、σ^2 未知的情况下，关于 σ^2 的检验问题.
  - 双边 χ^2 检验法
  - 单边 χ^2 检验法

化肥厂例子

某车间用一台自动包装机装化肥，每袋的标称重量规定为某日开工后为了检查包装机是否正常，随机抽检了 9 袋产品，测得净重如下

$$49.6,\ 48.4,\ 50.0,\ 48.6,\ 50.6,\ 49.8,\ 50.2,\ 49.2,\ 49.0.$$

设每袋化肥的实际重量服从正态分布，其标准差为 0.75 (kg). 试问该日包装机工作是否正常？

分析

设每袋化肥重量为 $X \sim \mathcal{N}(\mu, \sigma_0^2)$, $\sigma_0 = 0.75$. 若 $\mu = 50$ 则机器正常，否则不正常. μ 的点估计值 $\hat{\mu} = \bar{x} = 49.49 \neq 50$.

问: $\bar{x} \neq 50$ 是由于机器不正常还是随机波动引起?

- 提出假设: $H_0 : \mu = \mu_0 = 50$, $H_1 : \mu \neq \mu_0$
- 作出判断:
  - 拒绝 H_0, 即认为机器不正常
  - 接受 H_0, 即认为机器正常
- 作出判断的过程称为对假设 H_0 进行检验.

怎样检验假设?

分析

$$H_0: \mu = \mu_0 = 50, \quad H_1: \mu \neq \mu_0.$$

$X_1, \dots, X_n$ 是总体 $X \sim \mathcal{N}(\mu, \sigma_0^2)$ 的样本，且

$$\mathbb{E}\bar{X} = \mu, \quad \mathrm{Var}(\bar{X}) = \frac{\sigma_0^2}{n}.$$

故若 H_0 成立，则 $\mathbb{E}\bar{X} = \mu_0$，即 $\bar{X}$ 的值在 μ_0 附近波动，且波动幅度不太大，从而绝对偏差 $|\bar{X} - \mu_0|$ 应该较小.
反之，若 $|\bar{X} - \mu_0|$ 的值偏大，则有理由怀疑 H_0 的正确性从而拒绝 H_0.

可以求一临界值 C，使得

- 当 $|\bar{X} - \mu_0| \geq C$ 时，拒绝 H_0，
- 当 $|\bar{X} - \mu_0| < C$ 时，接受 H_0.

某车间用一台自动包装机装化肥，每袋的标称重量规定为某日开工后为了检查包装机是否正常，随机抽检了 9 袋产品，测得净重如下

$$49.6,\ 48.4,\ 50.0,\ 48.6,\ 50.6,\ 49.8,\ 50.2,\ 49.2,\ 49.0.$$

设每袋化肥的实际重量服从正态分布，其标准差为 0.75 (kg). 试问该日包装机工作是否正常？

检验方法

$$H_0: \mu = \mu_0 = 50 \text{ (原假设)}, \quad H_1: \mu \neq \mu_0 \text{ (备择假设)}.$$

取 $\alpha = 0.05$ 使得 I 类错误

$$\mathsf{P}_{H_0}\left(\frac{|\bar{X}-\mu_0|}{\sigma_0/\sqrt{n}} > \frac{C}{\sigma_0/\sqrt{n}}\right) = \alpha.$$

当 H_0 成立时，(检验统计量)$U = \dfrac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} \sim \mathcal{N}(0,1)$，所以 $\dfrac{C}{\sqrt{\sigma_0/\sqrt{n}}} = u_{1-\frac{\alpha}{2}} = 1.96$，

即当 $|U| > 1.96$ (拒绝域) 时拒绝 H_0. 具体计算，得 $|U| = \dfrac{|49.49-50|}{0.75/\sqrt{9}} = 2.04 > 1.96$，

故拒绝 H_0，即认为包装机工作不正常.

结果分析

假设

$$H_0: \mu = \mu_0 = 50, \quad H_1: \mu \neq \mu_0$$

的提出基于：

- 当日开工时机器并未出现明显异常
- 检验者认为机器正常，反映了生产车间的倾向性

因为 α 较小，且

$$\mathrm{P}\left(\frac{|\bar{X}-\mu_0|}{\sigma_0/\sqrt{n}} > u_{1-\alpha/2} \mid H_0\right) = \alpha,$$

说明当 H_0 为真时，$\left\{\frac{|\bar{X}-\mu_0|}{\sigma_0/\sqrt{n}} > u_{1-\alpha/2}\right\}$ 是小概率事件。而根据实际推断原理，如果该事件发生了，当然要怀疑 H_0 的真实性，从而作出拒绝 H_0 的决策。

检验原则三：概率反证法

如果想检验某“结论”成立，则应提出原假设：

$$H_0: \text{“结论”不成立.}$$

小结

检验的原则

- 检验原则一 (基本原则)：对 H_0 采取保护的态度
- 检验原则二：控制 I 类风险 (用于构造拒绝域)
- 检验原则三：概率反证法 (提出原假设)

依据上述原则，对原假设 H_0 作出“拒绝”或“接受”的决策，称为在显著性水平 α 下，对 H_0 进行**显著性检验**.

注

拒绝 H_0 即意味着接受 H_1. 故 H_0 的拒绝域即是 H_1 的接受域.

例

从甲地发送一个讯号到乙地，由于存在线路噪声干扰，使得甲地发送一个幅值为 μ 的讯号，而乙地收到的讯号是一个服从 $\mathcal{N}(\mu,4)$ 分布的随机变量．在测试中，甲地将同一讯号发送了 5 次，乙地收到的讯号值为

$$8.1,\ 9.3,\ 9.9,\ 8.5,\ 10.1.$$

接收方有某种理由猜测甲地发送的讯号值为 8，问这种猜测是否正确？($\alpha=0.05$)

解

依题意，要检验

$$H_0:\mu=\mu_0=4,\quad H_1:\mu\neq\mu_0.$$

采用 u 检验法，求得 H_0 的拒绝域为

$$|\bar{X}-\mu_0|>\frac{\sigma_0}{\sqrt{n}}u_{1-\alpha/2}=\frac{2}{\sqrt{5}}\times u_{0.975}=1.75.$$

又算得 $\bar{x}=9.18$．因为 $|\bar{x}-\mu_0|=|9.18-8|=1.18<1.75$，故不能拒绝 H_0，即可认为接收方的猜测是有道理的 (95% 可能是正确的)．

处理假设检验问题的一般方法和步骤

1. 根据实际问题，提出原假设 H_0 及备择假设 H_1.
2. 求出未知参数的较好的点估计
3. 依据点估计构造一个检验统计量，然后分析当成立时，该统计量有什么“趋势”，逆这个“趋势”就给出了 H_0 的拒绝域形式，即 H_0 的拒绝域形式由 H_1 确定
4. 对于给定的显著性水平 α，按控制I类风险的检验原则，确定 H_0 的拒绝域.
5. 抽样，判断样本观察值是否落在拒绝域内，从而作出“拒绝”或“接受(不拒绝)”的决策

例

某厂生产的固体燃料推进器燃烧率服从正态分布 $\mathcal{N}(40, 2^2)$ (cm/s). 现用新方法生产了一批推进器，从中随机取 $n = 25$ 只，试验后算得 $\bar{x} = 41.25$. 设新方法的总体方差不变. 问新方法燃烧率是否有显著提高？($\alpha = 0.05$)

分析

- **有什么理由提出“新方法燃烧率有显著提高”?**
  原平均燃烧率为 40, 而新方法燃烧率样本均值为 $\bar{x} = 41.25 > 40$.
- **依题意，要检验的假设是:**

$$H_0 : \mu \leq \mu_0 = 40, \quad H_1 : \mu > \mu_0.$$

  若采用 $H_0 : \mu = \mu_0 = 40$, $H_1 : \mu \neq \mu_0$ 或 $H_0 : \mu \geq \mu_0 = 40, H_1 : \mu < \mu_0$, 则变成了**保护“新方法”**. 我们更希望**“维持现状”**.
- **检验的目的是什么?**
  ~~检验的目的是要论证新方法好.~~ 检验的目的是要对新方法进行鉴定. 为了慎重起见，防止误判，我们宁可相信新方法的燃烧率没有提高.

例

某厂生产的固体燃料推进器燃烧率服从正态分布 $\mathcal{N}(40, 2^2)$ (cm/s). 现用新方法生产了一批推进器，从中随机取 $n = 25$ 只，试验后算得 $\bar{x} = 41.25$. 设新方法的总体方差不变. 问新方法燃烧率是否有显著提高? $(\alpha = 0.05)$

设新方法燃烧率 $X \sim \mathcal{N}(\mu, \sigma_0^2)$, 依题意, 要进行**单边检验**

$$H_0 : \mu \le \mu_0 = 40, \quad H_1 : \mu > \mu_0.$$

$\bar{X}$ 是 μ 的无偏估计, 且 $\mathrm{Var}(\bar{X}) = \dfrac{\sigma_0^2}{n}$. $\bar{X}$ 的值应在 μ 附近波动, , 且当 H_0 为真时, 统计量 $\bar{X} - \mu_0$ 的值应该偏小, 否则有理由怀疑 H_0 的真实性, 从而拒绝 H_0.

如何构造拒绝域?

$$\mathsf{P}\left(\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} > u_{1-\alpha} \mid H_0\right) \le \mathsf{P}\left(\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} > u_{1-\alpha} \mid \mu = \mu_0\right) = \alpha.$$

如果 $\dfrac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} > u_{1-\alpha}$ 作为 H_0 的拒绝域, 其 I 类风险不超过 α. μ 和 μ_0 距离越大越容易分辨, 越接近越难分辨, 重合时 I 类风险最大.

例

某厂生产的固体燃料推进器燃烧率服从正态分布 $\mathcal{N}(40, 2^2)$ (cm/s). 现用新方法生产了一批推进器, 从中随机取 $n = 25$ 只, 试验后算得 $\bar{x} = 41.25$. 设新方法的总体方差不变. 问新方法燃烧率是否有显著提高? ($\alpha = 0.05$)

解

设新方法燃烧率 $X \sim \mathcal{N}(\mu, \sigma_0^2)$, 依题意, 要进行**单边检验**

$$H_0 : \mu \leq \mu_0 = 40, \quad H_1 : \mu > \mu_0.$$

$\bar{X}$ 是 μ 的无偏估计, 且 $\mathrm{Var}(\bar{X}) = \dfrac{\sigma_0^2}{n}$. $\bar{X}$ 的值应在 μ 附近波动, , 且当 H_0 为真时, 统计量 $\bar{X} - \mu_0$ 的值应该偏小, 否则有理由怀疑 H_0 的真实性, 从而拒绝 H_0. 又当 $\mu = \mu_0$ 时, 有

$$\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

(单边假设检验问题, 检验统计量都采用临界点处的分布.) 故对于给定的显著水平 $\alpha = 0.05$, 求得 H_0 的拒绝域为 $\bar{X} - \mu_0 > \dfrac{\sigma_0}{\sqrt{n}} u_{1-\alpha} = \dfrac{2}{\sqrt{25}} \times 1.65 = 0.66$. 而 $\bar{x} - \mu_0 = 1.25 > 0.66$, 所以我们拒绝 H_0, 即认为新方法燃烧率有显著提高.

假设检验步骤总结

提出假设 根据统计调查的目的，提出原假设 H_0 和备选假设 H_1.

抽取样本 简单随机抽样

检验假设 对差异进行定量的分析，确定其性质(是随机误差还是系统误差. 为给出两者界限，找一检验统计量 T，在 H_0 成立下其分布已知.)

- 显著性水平 α: $\mathsf{P}(T \in W) = \alpha$, α 为犯第 I 类错误的概率，W 为拒绝域.

作出决策 拒绝还是接受 H_0.

作业

1. 某电器元件平均电阻值一直保持在 2.64 Ω，今测得采用新工艺生产 36 个元件的平均阻值为 2.61 Ω. 假定在正常条件下，电阻值服从正态分布，而且新工艺不改变电阻的标准差. 已知改变工艺前的标准偏差为 0.06 Ω. 问新工艺对产品的电阻值是否有显著性影响 $(\alpha = 0.01)$?
2. 某厂生产的某种钢索断裂强度服从正态 $\mathcal{N}(\mu, \sigma^2)$，其中 $\sigma = 40(\mathrm{kg/cm}^2)$. 现在一批这种钢索的容量为 9 的一个样本测得断裂强度平均值为 $\bar{X}$，与以往正常生产的 μ 相比，$\bar{X}$ 较 μ 大 $20(\mathrm{kg/cm}^2)$. 设总体方差不变，问在 $\alpha = 0.01$ 的显著性水平下，能否认为这批钢索质量显著提高?
3. 某织物强力指标 X 的均值 $\mu_0 = 21$ 公斤. 改进工艺后生产一批织物，今从中取 30 件，测得 $\bar{x} = 21.55$ 公斤. 假设强力指标服从正态分布 $\mathcal{N}(\mu, \sigma^2)$，且已知有 $\sigma = 1.2$ 公斤. 问在显著性水平 $\alpha = 0.01$ 下，新生产织物比过去的织物强力是否有提高?
4. 某种零件的尺寸方差为 $\sigma^2 = 1.21$. 对这一批这类零件检查 6 件得尺寸数据 (mm): 32.56, 29.66, 31.64, 30.00, 21.87, 31.03. 设零件尺寸服从正态分布，问这批零件的平均尺寸是否认为是 32.50 mm $(\alpha = 0.05)$?

1 概述: 基本思想

2 正态总体参数的假设检验

- 单总体 $\mathcal{N}(\mu, \sigma^2)$
- 双总体 $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mathcal{N}(\mu_2, \sigma_2^2)$
- 假设检验问题讨论
- 作业

回顾: 处理假设检验问题的一般方法和步骤

1. 根据实际问题, 提出原假设 H_0 及备择假设 H_1.
2. 求出未知参数的较好的点估计
3. 依据点估计构造一个检验统计量, 然后分析当成立时, 该统计量有什么 "趋势", 逆这个 "趋势"就给出了 H_0 的拒绝域形式, 即 H_0 的拒绝域形式由 H_1 确定
4. 对于给定的显著性水平 α , 按控制I 类风险的检验原则, 确定 H_0 的拒绝域.
5. 抽样, 判断样本观察值是否落在拒绝域内, 从而作出 "拒绝" 或 "接受 (不拒绝)" 的决策

单边假设检验问题, 检验统计量都采用临界点处的分布.

小结

检验的原则

- 检验原则一 (基本原则): 对 H_0 采取保护的态度
- 检验原则二: 控制 I 类风险 (用于构造拒绝域)
- 检验原则三: 概率反证法 (提出原假设)

u 检验法

单正态总体在 σ_0^2 已知的情况下，关于 μ 的检验问题

- 双边 u 检验法
- 单边 u 检验法

本节将讨论单、双正态总体的其他检验法

- t 检验法
- χ^2 检验法
- F 检验法

双边检验问题与双边 t 检验法

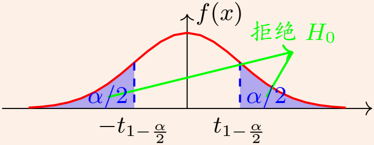
问题

设 $X_1,\dots,X_n$ 是总体 $X\sim\mathcal{N}(\mu,\sigma^2)$ 的样本，μ, σ^2 均未知，试在显著性水平 α 下，检验假设

$$H_0:\mu=\mu_0,\quad H_1:\mu\neq\mu_0\quad(\mu_0\text{ 已知}).$$

$\bar{X}$, S^2 分别是 μ, σ^2 的无偏估计，故当 H_0 为真时，统计量 $\dfrac{|\bar{X}-\mu_0|}{S/\sqrt{n}}$ 的值应该偏小，否则便要拒绝 H_0. 又当 H_0 为真时，

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



故由 $\mathrm{P}\left(\dfrac{|\bar{X}-\mu_0|}{S/\sqrt{n}}>t_{1-\alpha/2}(n-1)\mid\mu=\mu_0\right)=\alpha$，求得 H_0 的拒绝域为

$|\bar{X}-\mu_0|>\dfrac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)$ **(双边 t 检验法)**.

单边检验问题与单边 t 检验法

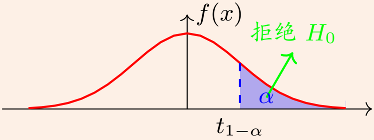
问题

设 $X_1, \ldots, X_n$ 是总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本，μ, σ^2 均未知，试在显著性水平 α 下，检验假设

$$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0 \quad (\mu_0 \text{ 已知}).$$

$\bar{X}$, S^2 分别是 μ, σ^2 的无偏估计，故当 H_0 为真时，统计量 $\frac{|\bar{X} - \mu_0|}{S/\sqrt{n}}$ 的值应该偏小于 0，否则便要拒绝 H_0. 在临界点 $\mu = \mu_0$ 处有

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



故由 $\mathrm{P}\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{1-\alpha}(n-1) \mid \mu = \mu_0\right) = \alpha$，求得 H_0 的拒绝域为

$\bar{X} - \mu_0 > \frac{S}{\sqrt{n}} t_{1-\alpha}(n-1)$ **(单边 t 检验法)**.

例题

某种元件的寿命 $X\sim\mathcal{N}(\mu,\sigma^2)$, μ, σ^2 均未知. 现测得 16 只元件的寿命 (小时) 如下：

$$159,\ 280,\ 101,\ 212,\ 224,\ 379,\ 179,\ 264,\ 222,\ 362,\ 168,\ 250,\ 149,\ 260,\ 485,\ 170.$$

问能否认为元件的平均寿命大于 225(小时)?

分析

- 为什么会提出 "平均寿命大于 225 小时" 的问题?
  1. 在样本观察值中, 有 7 个数据值远大于 225, 有 3 个数据值接近 225
  2. $\bar{x}=241>225$ (貌似元件平均寿命大于 225)
  3. 实际问题需要
- 怎样提出假设?

  $H_0:\mu=\mu_0=225,\quad H_1:\mu\neq\mu_0.$

  (正确) $H_0:\mu\le\mu_0=225,\quad H_1:\mu>\mu_0.$

  $H_0:\mu\ge\mu_0=225,\quad H_1:\mu<\mu_0.$

依题意，要检验假设 $H_0:\mu\le\mu_0=225, H_1:\mu>\mu_0$. 采用单边检验法, 求得 H_0 的拒绝域是 $X>\mu_0+\dfrac{S}{\sqrt{n}}t_{1-\alpha}(n-1)=225+43.269=268.269$, 而 $\bar{x}=241<268.269$. 故不拒绝 H_0, 即认为元件的平均寿命不大于 225(小时).

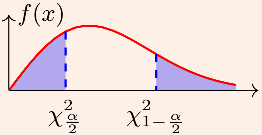
双边 χ^2 检验法

设 $X_1, \ldots, X_n$ 是总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, μ, σ^2 均未知, 试在显著性水平 α 下, 检验假设

$$H_0: \sigma^2 = \sigma_0^2, \quad H_1: \sigma^2 \neq \sigma_0^2 \quad (\sigma_0^2 \text{ 已知}).$$

S^2 分别是 σ^2 的无偏估计, 且 $\mathrm{Var}(S^2) = \dfrac{2\sigma^4}{n-1} \to 0$, , 故当 H_0 成立时统计量 S^2 的值应在 σ_0^2 附近波动, , 且幅度不应太大, 否则有理由拒绝 H_0. 又当 H_0 为真时,

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



故在显著性水平 α 下, H_0 的拒绝域为 $\left[0, \dfrac{\sigma_0^2}{n-1}\chi^2_{\frac{\alpha}{2}}(n-1)\right) \cup \left(\dfrac{\sigma_0^2}{n-1}\chi^2_{1-\frac{\alpha}{2}}(n-1), \infty\right)$.

若 μ 已知, 则 H_0 的检验统计量与分布为 $\dfrac{1}{\sigma_0^2}\displaystyle\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$.

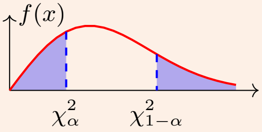
单边 χ^2 检验法

设 $X_1,\ldots,X_n$ 是总体 $X\sim\mathcal{N}(\mu,\sigma^2)$ 的样本, μ, σ^2 均未知, 试在显著性水平 α 下, 检验假设

$$H_0:\sigma^2\geq\sigma_0^2,\quad H_1:\sigma^2<\sigma_0^2\quad(\sigma_0^2\text{ 已知}).$$

S^2 分别是 σ^2 的无偏估计, 且 $\mathrm{Var}(S^2)=\dfrac{2\sigma^4}{n-1}\to 0$, , 故当 H_0 成立时统计量 S^2 的值应偏大于 σ_0^2, 若过于偏小于 σ_0^2, 则有理由拒绝 H_0. 又当 $\sigma^2=\sigma_0^2$ 为真时,

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



故在显著性水平 α 下, H_0 的拒绝域为 $\dfrac{(n-1)S^2}{\sigma_0^2}<\chi^2_\alpha(n-1)$.

单边检验问题 $H_0:\sigma^2\leq\sigma_0^2$, $H_1:\sigma^2>\sigma_0^2$ 的拒绝域是 $\dfrac{(n-1)S^2}{\sigma_0^2}>\chi^2_{1-\alpha}(n-1)$.

例

一台自动装配磁带机器装配每盒磁带长度服从正态分布，如果磁带长度的标准差不超过 0.15 厘米，认为机器正常，否则就要调整机器。现抽取 10 盒磁带，经检测算得样本方差观察值 $s^2 = 0.028(\mathrm{cm}^2)$. 试问机器是否正常？($\alpha = 0.05$)

解

依题意，要检验假设

$$H_0 : \sigma^2 \leq \sigma_0^2 = 0.0225, \quad H_1 : \sigma^2 > \sigma_0^2.$$

采用单边 χ^2 检验法，求得 H_0 的拒绝域是

$$\frac{(n-1)S^2}{\sigma_0^2} > \chi^2_{1-\alpha}(n-1).$$

查表得 $\chi^2_{0.95}(9) = 16.919$，由于

$$(n-1)s^2/\sigma_0^2 = \frac{9 \times 0.028}{0.0225} = 11.2 < 16.929,$$

故不拒绝 H_0，即可认为机器工作正常.

单个正态总体参数的假设检验使用的检验统计量

检验均值 (方差已知)

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

检验均值（方差未知）

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

检验方差（均值已知）

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

检验方差（均值未知）

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

关于 μ 的检验法

u 检验法 (σ^2 已知)

检验统计量及 H_0 时的分布	原假设 H_0	备择假设 H_1	拒绝域
$U = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1)$	$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert U \rvert \geq u_{1-\frac{\alpha}{2}}$
	$\mu \geq \mu_0$	$\mu < \mu_0$	$U \leq u_\alpha$
	$\mu \leq \mu_0$	$\mu > \mu_0$	$U \geq u_{1-\alpha}$

t 检验法 (σ^2 未知)

检验统计量及 H_0 时的分布	原假设 H_0	备择假设 H_1	拒绝域
$T = \dfrac{\bar{X} - \mu_0}{S / \sqrt{n}} \sim t(n-1)$	$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert T \rvert \geq t_{1-\frac{\alpha}{2}}$
	$\mu \geq \mu_0$	$\mu < \mu_0$	$T \leq t_\alpha$
	$\mu \leq \mu_0$	$\mu > \mu_0$	$T \geq t_{1-\alpha}$

关于 σ^2 的检验法

χ^2 检验法 (μ 已知)

检验统计量及 H_0 时的分布	原假设 H_0	备择假设 H_1	拒绝域
$\chi^2=\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2}\sim\chi^2(n)$	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2\notin[\chi^2_{\frac{\alpha}{2}}(n),\chi^2_{1-\frac{\alpha}{2}}(n)]$
	$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$	$\chi^2\leq\chi^2_{\alpha}(n)$
	$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$	$\chi^2\geq\chi^2_{1-\alpha}(n)$

χ^2 检验法 (μ 未知)

$(n'=n-1)$

检验统计量及 H_0 时的分布	原假设 H_0	备择假设 H_1	拒绝域
$\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n')$	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2\notin[\chi^2_{\frac{\alpha}{2}}(n'),\chi^2_{1-\frac{\alpha}{2}}(n')]$
	$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$	$\chi^2\leq\chi^2_{\alpha}(n')$
	$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$	$\chi^2\geq\chi^2_{1-\alpha}(n')$

问题背景

设研究对象的某指标 $X \sim \mathcal{N}(\cdot, \cdot)$. 如果外界条件发生了变化，则要研究外界条件的变化是否对该指标产生了影响.

分析

设变化前指标 $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, 变化后指标 $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

$\mu_1 - \mu_2$ 的假设检验问题

1. $H_0 : \mu_1 - \mu_2 = \delta,\ H_1 : \mu_1 - \mu_2 \neq \delta.$
2. $H_0 : \mu_1 - \mu_2 \geq \delta,\ H_1 : \mu_1 - \mu_2 < \delta.$
3. $H_0 : \mu_1 - \mu_2 \leq \delta,\ H_1 : \mu_1 - \mu_2 > \delta.$

其中 δ 为已知常数, 通常 $\delta = 0$.

设 $X_1, \ldots, X_{n_1}$ 为总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, $Y_1, \ldots, Y_{n_2}$ 为总体 $Y \sim \mathcal{N}(\mu_2, \sigma^2)$ 的样本, 两样本独立, 试检验假设

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

解

μ_1, μ_2, σ^2 的无偏估计分别为

$$\bar{X}, \quad \bar{Y}, \quad S_\omega^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

其中 $\bar{X}, \bar{Y}$, S_1^2, S_2^2 分别为总体 X, Y 的样本均值和样本方差. 当 H_0 为真时, 统计量

$$T = \frac{\bar{X} - \bar{Y}}{S_\omega \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

且 $|T|$ 应该偏小, 如果偏大, 则应拒绝 H_0. 对于给定的显著性水平 α, H_0 的拒绝域是

$$\frac{|\bar{X} - \bar{Y}|}{S_\omega \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2).$$

设 $X_1, \ldots, X_{n_1}$ 为总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, $Y_1, \ldots, Y_{n_2}$ 为总体 $Y \sim \mathcal{N}(\mu_2, \sigma^2)$ 的样本, 两样本独立, 试检验假设

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

注

1. 若总体的方差均未知, 则 t 检验要求 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ **(方差齐性)**.
2. 若总体的方差 σ_1^2, σ_2^2 均已知, 则检验统计量可取为

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

   对应的检验为 u 检验.
3. 若总体方差及是否具有方差齐性均未知, 怎么办?
   可先检验假设

$$H_0 : \sigma_1^2 = \sigma_2^2, \quad H_1 : \sigma_1^2 \neq \sigma_2^2.$$

   若接受 H_0 则可再进行 t 检验.

为了提高化学产品得率, 采用了某新工艺. 在对比试验中, 分别用两种方法各独立进行了 8 次试验, 计算出

$$\text{新工艺: } \bar{x}=91.73,\ s_1^2=3.89, \quad \text{老工艺: } \bar{y}=93.75,\ s_2^2=4.02.$$

假定老、新工艺的得率分别为 $X\sim\mathcal{N}(\mu_1,\sigma^2)$, $Y\sim\mathcal{N}(\mu_2,\sigma^2)$. 试问新工艺是否能显著提高产品得率? $(\alpha=0.05)$

分析

由题给数据知 $\bar{x}<\bar{y}$. 有两种可能原因: 新工艺的确比老工艺好, 或新工艺不一定比老工艺好, 随机误差引起. 因此我们需要进行假设检验 (偏向维持现状):

$$H_0:\mu_1\geq\mu_2, \quad H_1:\mu_1<\mu_2.$$

解

采用单边 t 检验法，求得 H_0 的拒绝域为

$$\bar{X}-\bar{Y}<-t_{1-\alpha}(n_1+n_2-2)S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}=1.7613\times1.99\times\sqrt{\frac{1}{8}+\frac{1}{8}}=-1.752.$$

但 $\bar{x}-\bar{y}=-2.02<-1.752$. 因此拒绝 H_0, 即新工艺能显著地提高产品得率.

两总体方差的假设检验问题

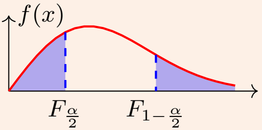
设 $X_1, \dots, X_{n_1}$ 为总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, $Y_1, \dots, Y_{n_2}$ 为总体 $Y \sim \mathcal{N}(\mu_2, \sigma^2)$ 的样本, 两样本独立, $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 均未知. 试在水平 α 下检验假设

$$H_0 : \sigma_1 = \sigma_2, \quad H_1 : \sigma_1 \neq \sigma_2.$$

解

两样本方差 S_1^2, S_2^2 分别是 σ_1^2, σ_2^2 的无偏估计. 故当 H_0 为真时, 统计量 S_1^2/S_2^2 的值应在常数 1 附近波动. 若 S_1^2/S_2^2 偏离 1 太远, 则有理由拒绝 H_0. 又当 H_0 成立时,

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



对于给定的显著性水平 α, H_0 的拒绝域是**(双边 F 检验法)**

$$\frac{S_1^2}{S_2^2} \in \left[0, F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)\right) \cup \left(F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1), \infty\right).$$

两总体方差的假设检验问题 II

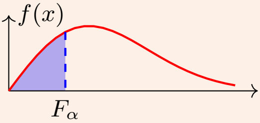
设 $X_1, \ldots, X_{n_1}$ 为总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, $Y_1, \ldots, Y_{n_2}$ 为总体 $Y \sim \mathcal{N}(\mu_2, \sigma^2)$ 的样本, 两样本独立, $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 均未知. 试在水平 α 下检验假设

$$H_0 : \sigma_1 \geq \sigma_2, \quad H_1 : \sigma_1 < \sigma_2.$$

解

两样本方差 S_1^2, S_2^2 分别是 σ_1^2, σ_2^2 的无偏估计. 故当 H_0 为真时, 统计量 S_1^2/S_2^2 的值应偏大于常数 1. 若 S_1^2/S_2^2 过于偏小于 1, 则有理由拒绝 H_0. 又当 $\sigma_1^2 = \sigma_2^2$ 成立时,

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



对于给定的显著性水平 α, H_0 的拒绝域是**(单边 F 检验法)**

$$\frac{S_1^2}{S_2^2} < F_\alpha(n_1 - 1, n_2 - 1).$$

例

一台机床大修前曾加工了 $n_1 = 10$ 件零件, 加工尺寸的样本方差为 $s_1^2 = 2500$, 大修后加工了 $n_2 = 12$ 件零件, 加工尺寸的样本方差为 $s_2^2 = 400$. 试问机床大修后的加工精度是否有显著提高? $(\alpha = 0.01)$.

解

大修前 $s_1^2 = 2500$, 大修后 $s_2^2 = 400$. 因为 $s_1^2 > s_2^2$, 貌似机床大修后加工精度有了提高. 我们采用概率反证法, 所提原假设与"貌似结论"相反, 提出如下假设

$$H_0 : \sigma_1^2 \le \sigma_2^2, \quad H_1 : \sigma_1^2 > \sigma_2^2.$$

检验目的是对机床大修后的加工精度进行验收, 保护委托方的利益.
采用单边 F 检验法, 求得 H_0 的拒绝域是

$$\frac{S_1^2}{S_2^2} > F_{1-\alpha}(n_1 - 1, n_2 - 1) = F_{0.99}(9, 11) = 4.63,$$

而 $\dfrac{s_1^2}{s_2^2} = \dfrac{2500}{400} = 6.25 > 4.63$, 故拒绝 H_0, 即认为大修后机床加工精度有显著提高.

双正态总体的检验法

$\mu_1-\mu_2$ 的检验 (σ_1^2,σ_2^2 已知)

检验统计量及 H_0 时的分布	原假设 H_0	备择假设 H_1	拒绝域
$U=\dfrac{\bar X-\bar Y-\delta}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}\sim\mathcal{N}(0,1)$	$\mu_1-\mu_2=\delta$	$\mu_1-\mu_2\neq\delta$	$\lvert U\rvert\geq u_{1-\frac{\alpha}{2}}$
	$\mu_1-\mu_2\geq\delta$	$\mu_1-\mu_2<\delta$	$U\leq u_\alpha$
	$\mu_1-\mu_2\leq\delta$	$\mu_1-\mu_2>\delta$	$U\geq u_{1-\alpha}$

$\mu_1-\mu_2$ 的检验 ($\sigma_1^2=\sigma_2^2=\sigma^2$, σ^2 未知)

检验统计量及 H_0 时的分布	原假设 H_0	备择假设 H_1	拒绝域
$T=\dfrac{\bar X-\bar Y-\delta}{\sqrt{\frac{1}{n}+\frac{1}{m}}S_\omega}\sim t(n+m-2)$	$\mu_1-\mu_2=\delta$	$\mu_1-\mu_2\neq\delta$	$\lvert T\rvert\geq t_{1-\frac{\alpha}{2}}$
	$\mu_1-\mu_2\geq\delta$	$\mu_1-\mu_2<\delta$	$T\leq t_\alpha$
	$\mu_1-\mu_2\leq\delta$	$\mu_1-\mu_2>\delta$	$T\geq t_{1-\alpha}$

$$S_\omega=\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}.$$

双正态总体的检验法 II

方差比 σ_1^2/σ_2^2 的检验 (μ_1, μ_2 已知)

$n' = n - 1, m' = m - 1$

检验统计量及 H_0 时的分布	原假设 H_0	备择假设 H_1	拒绝域
$F = \dfrac{S_1^2}{S_2^2} \sim F(n', m')$	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F \geq F_{1-\frac{\alpha}{2}}(n', m')$ 或 $F \leq F_{\frac{\alpha}{2}}(n', m')$
	$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F \leq F_{\alpha}(n', m')$
	$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F \geq F_{1-\alpha}(n', m')$

参数检验的一般提法

设总体 $X \sim F(x;\theta)$, $\theta \in \Theta$ 为未知参数，由实际问题需要，可作参数空间的分划 $\Theta = \Theta_0 \cup (\Theta \setminus \Theta_0)$, 则参数检验问题的一般提法是

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta - \Theta_0.$$

若存在 $W_0 \subset \mathbb{R}^n$, 有决策

$$\left.\begin{array}{ll} (X_1, \ldots, X_n) \in W_0, & \Rightarrow \text{拒绝 } H_0 \\ (X_1, \ldots, X_n) \notin W_0, & \Rightarrow \text{接受 } H_0 \end{array}\right\}$$

则称 W_0 为 H_0 的拒绝域，或称为 H_0 的一个检验.

风险

I 类风险 $\mathsf{P}\big((X_1,\ldots,X_n)\in W_0 \mid \theta\in\Theta_0\big)$

II 类风险 $\mathsf{P}\big((X_1,\ldots,X_n)\notin W_0 \mid \theta\in\Theta\setminus\Theta_0\big)$

显著性检验原则：控制 I 类风险 (不管 II 类风险)

给定显著性水平 α, $\alpha\in(0,1)$ 使得

$$\mathsf{P}\big((X_1,\ldots,X_n)\in W_0 \mid \theta\in\Theta_0\big)\le\alpha.$$

检验的优劣

若 W_1, W_2 都是 H_0 的显著性水平为 α 的检验, 问怎样评价与选择?
分析: W_1 与 W_2 的 I 类风险都不超过 α, 因此评价与选择的原则是 II 类风险较小的检验较优.

决策的可信度问题

对于假设问题

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta - \Theta_0.$$

设显著性水平为 α 的一个拒绝域为 W_0.

决策一 $(X_1, \ldots, X_n) \in W_0 \Rightarrow$ 拒绝 H_0.

决策二 $(X_1, \ldots, X_n) \notin W_0 \Rightarrow$ 接受 H_1.

问

决策一和决策二的哪个可信度高?

分析

- 决策一的 I 类风险受到控制, 其值不超过 α, 可信度高
- 决策二的 II 类风险有多大不清楚, 可信度未知

原假设的提出问题

检验的原则

原则一　对 H_0 采取保护的态度

原则二　控制 I 类风险

H_0, H_1 地位不对等!

- 某人去看医生，那么医生应该提出假设

$$H_0: \text{此人有病}, \quad H_1: \text{此人无病}.$$

- 鉴定某药品的疗效，应该提出假设

$$H_0: \text{此药无效}, \quad H_1: \text{此药有效}.$$

- 鉴定新技术应用效果，应该提出假设

$$H_0: \text{无效果}, \quad H_1: \text{有效果}.$$

- 在司法审判中，对犯罪嫌疑人应提出假设

$$H_0: \text{无罪}, \quad H_1: \text{有罪}.$$

1. 正常人的脉搏平均为 62 次/分. 今对某种疾病患者 10 人, 测其脉搏为 54, 68, 65, 77, 70, 64, 69, 72, 62, 71 次/分. 设患者的脉搏次数 X 服从正态分布. 试在显著性水平 $\alpha = 0.05$ 下, 检验患者的脉搏与正常人的脉搏有无差异?
2. 使用 A (电学法) 与 B (混合法) 两种方法来研究冰的潜热, 样品都是 $-0.72^\circ\mathrm{C}$ 的冰块, 下列数据是每克冰从 $-0.72^\circ\mathrm{C}$ 变成 $-0^\circ\mathrm{C}$ 水的过程中的吸热量 (卡/克):

$$
\begin{aligned}
\text{方法 A}: &79.98,\ 80.04,\ 80.02,\ 80.03,\ 80.03,\ 80.04,\ 80.04,\\
&79.97,\ 80.05,\ 80.03,\ 80.02,\ 80.00,\ 80.02,\\
\text{方法 B}: &80.02,\ 79.94,\ 79.97,\ 79.98,\ 79.97,\ 80.03,\ 79.95,\ 79.97.
\end{aligned}
$$

   假定每种方法测得的数据都服从正态分布, 且它们的方差相等. 检验 H_0 : 两种方法的总体均值相等 ($\alpha = 0.05$).
3. 加工某一机器零件, 根据其精度要求, 标准差不得超过 0.9. 现从该产品中抽取 19 个样本, 得样本标准差为 $S = 1.2$. 当 $\alpha = 0.05$ 时, 可否认为标准差变大? (假定零件尺寸服从正态分布.)
4. 从城市的某区中抽取 16 名学生测其智商, 平均值为 107, 样本标准差为 10. 而从该城市的另一区抽取的 16 名学生平均智商为 112, 标准差为 8, 试问在显著性水平 $\alpha = 0.05$ 下, 这两组学生的智商有无差异? 假定学生的智商服从正态分布.

5. 下表给出两个文学家 Mark Twain 的 8 篇小品文以及 Snodgrass 的 10 篇小品文中由 3 个字母组成的词的比例.
设两组数据分别来自正态总体, 且两总体方差相等, 两样本相互独立. 问两个作家所写的小品文中包含由 3 个字母组成的词的比例是否有显著性差异 ($\alpha = 0.05$)?

Mark Twain:$0.225,\ 0.262,\ 0.217,\ 0.240,\ 0.230,\ 0.229,\ 0.235,\ 0.217,$

Snodgrass$0.209,\ 0.205,\ 0.196,\ 0.210,\ 0.202,\ 0.207,\ 0.224,\ 0.223,\ 0.220,\ 0.201.$