

第三章：假设检验

(Hypothesis Tests)

何志坚

华南理工大学数学学院

hezhijian@scut.edu.cn



目录

1. 假设检验基本概念

- 1.1 检验问题的提出
- 1.2 检验法则与两类错误
- 1.3 检验水平与功效

2. 似然比检验

- 2.1 似然比检验法原理
- 2.2 Neyman-Pearson引理
- 2.3 广义似然比检验法原理

目录

3. 正态总体的检验

- 3.1 单个正态总体的假设检验
- 3.2 两个独立正态总体的假设检验

4. 拟合优度检验

- 4.1 卡方检验法
- 4.2 独立性检验
- 4.3 柯尔莫哥洛夫检验法
- 4.4 正态性检验

案例1：女生品茶

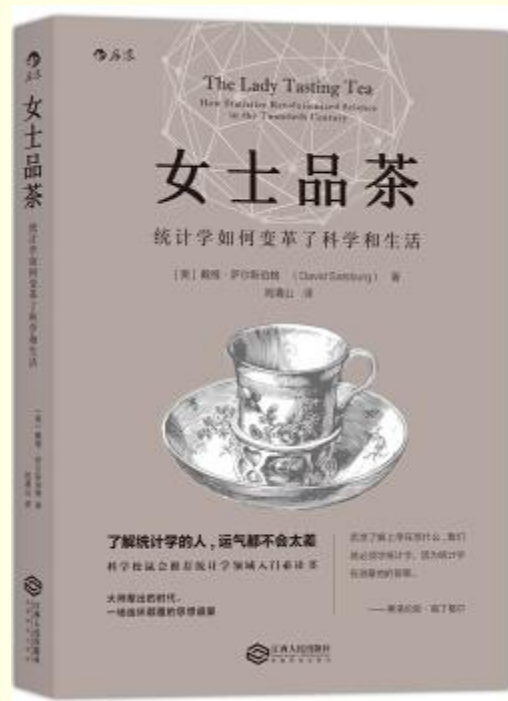
奶茶是由牛奶与茶按一定比例混合而成，可以先倒茶后加奶，也可以先倒奶再倒茶。某女士声称她可以鉴别这两种混合方式，周围品茶的人对此产生了议论，都觉得不可思议。在场的费希尔也在思考这个问题，他提议做一项试验来检验如下命题是否可以接受：

假设H: 该女士无此种鉴别能力

他准备了10杯调好的奶茶（两种顺序的都有）给该女士鉴别，**结果那位女士竟然能够正确地分辨出10杯奶茶中的每一杯的调制顺序**

。

如何做出你的判断？



案例2：鸢尾花

例 3.1 (Fisher's iris flower data set) Fisher的鸢尾花数据集是个著名的数据集，是为了量化鸢尾花形态上的区别而收集得到的。它包含了三类鸢尾花——山鸢尾(Setosa)、杂色鸢尾(Versicolour)、维吉尼亚鸢尾(Virginica)的花萼(sepal)、花瓣(petal)的长度与宽度。



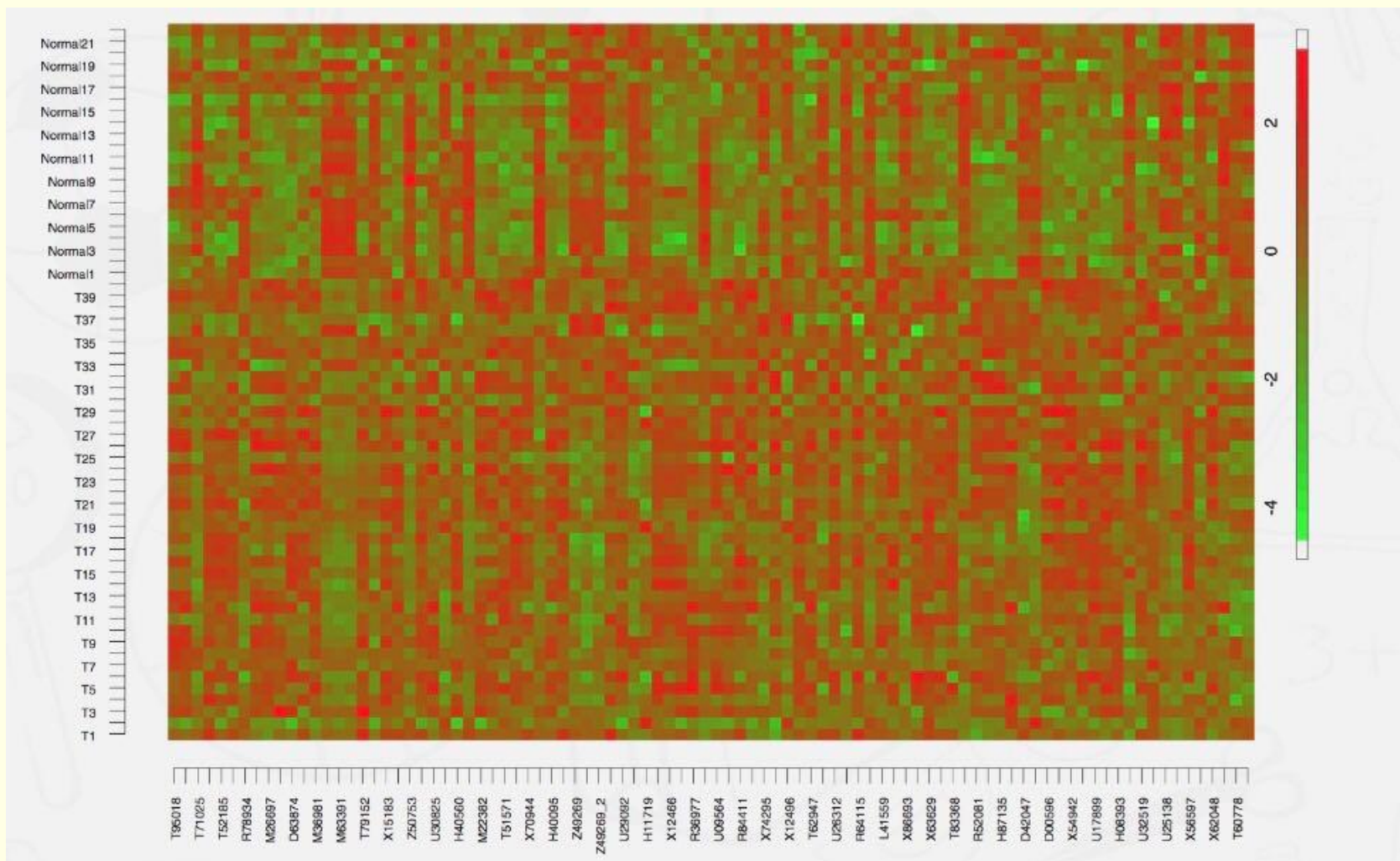
案例2：鸢尾花

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
51	7.0	3.2	4.7	1.4	versicolor
52	6.4	3.2	4.5	1.5	versicolor
53	6.9	3.1	4.9	1.5	versicolor
101	6.3	3.3	6.0	2.5	virginica
102	5.8	2.7	5.1	1.9	virginica
103	7.1	3.0	5.9	2.1	virginica

考虑以下几个问题：

1. 假设有个植物学家跟你说，通过基因组分析表明，山鸢尾(Setosa)花萼长度的均值是4.5cm，他这个论断是否可信？
2. 另一位植物学家说山鸢尾(Setosa)花萼长度的均值是一个不小于4.5cm的数，但具体是多少就不清楚了，那么他这个论断是否又可信？
3. 山鸢尾(Setosa)和杂色鸢尾(Versicolour)两种花的花萼长度的均值有没有显著差异？

案例3：基因表达水平比较



其它例子

1. 产品的次品率是否不超过3%
2. 男生群体平均身高是否大于女生群体平均身高?
3. 身高是否服从正态分布?
4. 抽烟与慢性支气管炎是否有关?

1. 基本概念

设有来自某一参数分布族 $\{F(x, \theta)\}$, $\theta \in \Theta$, 其中 Θ 为参数空间。

原假设（零假设, **Null Hypothesis**）

$$H_0: \theta \in \Theta_0$$

备选假设（对立假设, 备择假设, **Alternative Hypothesis**）

$$H_1: \theta \in \Theta_1$$

其中 $\emptyset \neq \Theta_0, \Theta_1 \subset \Theta, \Theta_0 \cap \Theta_1 = \emptyset$. 最常见的情况 $\Theta_1 = \Theta - \Theta_0$.

- 简单原假设(**simple null**): Θ_0 只包含一个点, 如 $H_0: \theta = \theta_0$
- 复杂原假设(**composite null**): Θ_0 只包含多个点, 如 $H_0: \theta \leq \theta_0$

通常有三种可能:

$H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$ 双边(**two-sided**)

$H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$ 单边(**one-sided**)

$H_0: \theta \geq \theta_0$ vs. $H_1: \theta < \theta_0$ 单边(**one-sided**)

决策规则

检验法则：检验本质上是把样本空间划分成两个互不相交的部分 W 和 \overline{W} , 当样本属于 W 时就拒绝 H_0 ; 否则接受 H_0 . 称 W 为该检验的拒绝域(rejection region), 而 \overline{W} 为接受域(acceptance region).

例：假设总体为 $N(\theta, 1)$, 样本为 X_1, \dots, X_n . 考虑检验问题

$$H_0: \theta = \theta_0 \text{ vs. } H_1: \theta \neq \theta_0$$

拒绝域可以设为: $W = \{(x_1, \dots, x_n): |\bar{x} - \theta_0| \geq c\}$

一个检验法则对应一种拒绝域; 一个拒绝域决定一种检验法则。

假设检验的目标是在给定准则下选取合适的拒绝域。

- 原假设 H_0 在客观上只有“真”和“假”两种可能
- 样本也只有两种可能性: $(X_1, \dots, X_n) \in W$ 或者 $(X_1, \dots, X_n) \notin W$
- 由于样本的随机性, 检验不可能100%正确

两类错误

客观事实	接受原假设	拒绝原假设
H_0 为真	正确	第一类错误(Type I error) 拒真错误 假阳性(false positive)
H_1 为真 (H_0 为假)	第二类错误(Type II error) 纳伪错误 假阴性(false negative)	正确

我们关心两类概率（条件概率）：

$$\alpha = P(\text{犯第一类错误}) = P((X_1, \dots, X_n) \in W | H_0)$$

$$\beta = P(\text{犯第二类错误}) = P((X_1, \dots, X_n) \notin W | H_1)$$

两类错误的关系

例：假设总体为 $N(\theta, 1)$, 样本为 X_1, \dots, X_n . 考虑简单的检验问题(simple hypothesis)

$$H_0: \theta = \theta_1 \text{ vs. } H_1: \theta = \theta_2$$

其中 $\theta_1 \neq \theta_2$.

选择拒绝域为 $W = \{(x_1, \dots, x_n): |\bar{x} - \theta_1| \geq c\}$.

- 犯第一类错误的概率为

$$\alpha = P_{\theta_1}(|\bar{X} - \theta_1| \geq c) = 2 - 2\Phi(c \sqrt{n})$$

- 犯第二类错误的概率为

$$\beta = P_{\theta_2}(|\bar{X} - \theta_1| < c) = \Phi((\theta_1 - \theta_2 + c) \sqrt{n}) - \Phi((\theta_1 - \theta_2 - c) \sqrt{n})$$

我们观察到：

- 如果 α 变小，则 c 增大，于是 β 变大。
- 如果 β 变小，则 c 减小，于是 α 变大。

结论：在样本量不变的前提下，两类错误不能同时减小。

显著性水平和功效

拒绝域的选取准则：在保证犯第一类错误的概率不超过一定水平的前提下，选择犯第二类错误的概率尽可能小的拒绝域 W

功效函数(power function):

$$\rho_W(\theta) := P_\theta((X_1, \dots, X_n) \in W)$$

- 当 $\theta \in \Theta_0$ 时， $\alpha = \rho_W(\theta)$ 表示犯第一类错误的概率
- 当 $\theta \in \Theta_1$ 时， $\beta = 1 - \rho_W(\theta)$ 表示犯第二类错误的概率， $\rho_W(\theta)$ 表示检验的功效(power of the test)

W 的检验水平（显著性水平，level of significance）：

$$\sup_{\theta \in \Theta_0} \rho_W(\theta) = \alpha$$

- 一般取 $\alpha = 0.1, 0.05, 0.01$

小概率原理：小概率事件在一次试验中是几乎不发生的。若 H_0 为真，样本落在拒绝域 W 是小概率事件，不应发生。如发生，则拒绝原假设。

如何选择显著性水平?

概率小到什么程度才当作“小概率事件”呢?

这要根据实际情况而定，例如即使下雨的概率为10%，仍有人会因为它太小而不带雨具。但某航空公司的事故率为1%，人们就会因为它太大而不敢乘坐该公司的飞机。

通常把概率不超过0.05 (或0.01)的事件当作“小概率事件 (rare event)”。为此在假设检验时，必须先确定小概率即显著性的值 α (即不超过 α 的概率认为是小概率)。

UMP检验

定义1: 称 W 为检验水平 α 的一致最大功效(uniformly most powerful, UMP)的拒绝域, 若 W 的检验水平为 α 且对一切检验水平不超过 α 的拒绝域 W' 均有

$$\rho_W(\theta) \geq \rho_{W'}(\theta), \forall \theta \in \Theta_1$$

- UMP意味着在犯第一类错误的概率不超过 α 的前提下, 犯第二类错误的概率最小(即功效最大)。
- 即不存在一种第一类错误和第二类错误都小于UMP检验的检验

Case study: two coins

Suppose that I have two coins, coin 0 has probability of heads equal to **0.5** and coin 1 has probability of heads equal to **0.7**. I choose one of the coins, **toss it 10 times** and tell you the number of heads, but do not tell you whether it was coin 0 or coin 1. On the basis of the number of heads, your task is to decide which coin it was. How should your decision rule be?

- let X be the number of heads
- let θ be the coin that produce this result.
- your test might be: $H_0: \theta = 0$ vs. $\theta = 1$

The probability mass function (PMF)

x	0	1	2	3	4	5	6	7	8	9	10
coin 0	.0010	.0098	.0439	.1172	.2051	.2461	.2051	.1172	.0439	.0098	.0010
coin 1	.0000	.0001	.0014	.0090	.0368	.1029	.2001	.2668	.2335	.1211	.0282

x	0	1	2	3	4	5	6	7	8	9	10
$\frac{P(x H_0)}{P(x H_1)}$	165.4	70.88	30.38	13.02	5.579	2.391	1.025	0.4392	0.1882	0.0807	0.0346

Suppose that you observed two heads. Then $P_0(2)/P_1(2) \approx 30$, which we will call the **likelihood ratio (LR)**—coin 0 was about 30 times more likely to produce this result than was coin 1. This result would favor coin 0.

On the other hand, if there were 8 heads, the likelihood ratio would be $P_0(8)/P_1(8) \approx 0.19$, which would favor coin 1.

The likelihood ratio will play a central role in the procedures we develop.

2. Likelihood ratio tests

考虑最简单的假设检验($\theta_1 \neq \theta_2$):

$$H_0: \theta = \theta_1 \text{ vs. } H_1: \theta = \theta_2$$

令 $\vec{x} = (x_1, \dots, x_n)$, 似然比定义为

$$LR = \frac{L(\vec{x}; \theta_2)}{L(\vec{x}; \theta_1)}.$$

似然比检验的拒绝域为: $W = \{\vec{x}: LR > \lambda\}$,

其中 $\lambda \geq 0$ 满足

$$P_{\theta_1}((X_1, \dots, X_n) \in W) = \int_W L(\vec{x}; \theta_1) d\vec{x} = \alpha$$

Neyman-Pearson Lemma(N-P引理): 上述定义的似然比检验 W 是一致最大功效的(UMP)

UMP tests for normal mean

例：假设总体为 $N(\mu, \sigma^2)$, 其中 σ^2 已知, 样本为 X_1, \dots, X_n . 考虑显著性水平为 α 的检验问题 ($\mu_2 > \mu_1$),

$$H_0: \mu = \mu_1 \text{ vs. } H_1: \mu = \mu_2$$

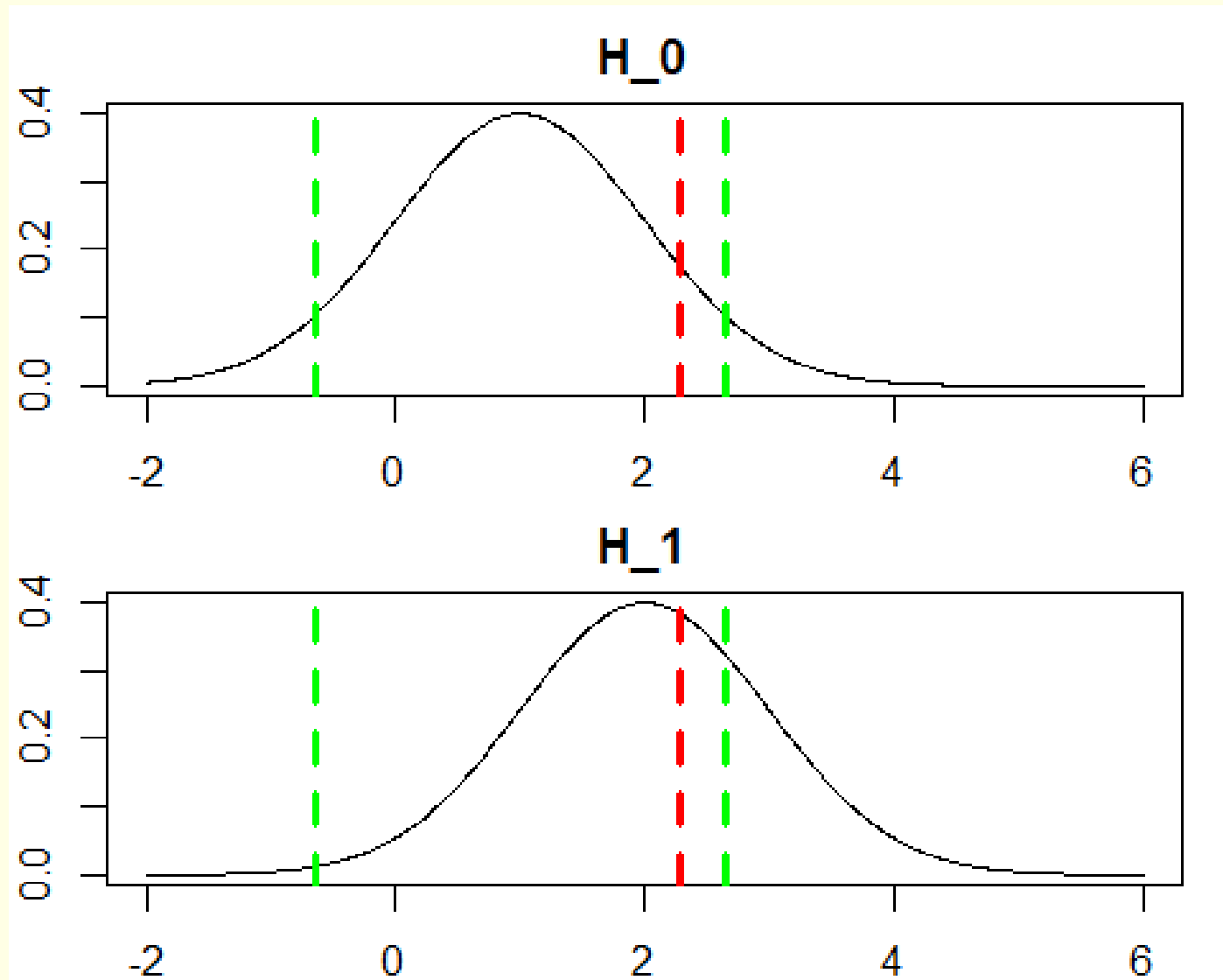
解：似然比检验的拒绝域为： $W = \{\vec{x} : \frac{L(\vec{x}; \mu_2)}{L(\vec{x}; \mu_1)} > \lambda\}$. 令似然比

$$\frac{L(X_1, \dots, X_n; \mu_2)}{L(X_1, \dots, X_n; \mu_1)} = \prod_{i=1}^n \frac{f(X_i; \mu_2, \sigma^2)}{f(X_i; \mu_1, \sigma^2)} = e^{\frac{n(\mu_2 - \mu_1)(2\bar{X} - \mu_1 - \mu_2)}{2\sigma^2}}$$

等价于找 C , 使得 $P_{\mu_1}(\bar{X} > C) = \alpha$, $C = \mu_1 + u_{1-\alpha}\sigma / \sqrt{n}$, 所以似然比检验的拒绝域为 $W = \{\vec{x} : \bar{x} > \mu_1 + u_{1-\alpha}\sigma / \sqrt{n}\}$

与双侧拒绝域比较： $W' = \{\vec{x} : |\bar{x} - \mu_1| > u_{1-\alpha/2}\sigma / \sqrt{n}\}$, 哪个好?

Comparison



UMP tests for normal mean

例：假设总体为 $N(\mu, \sigma^2)$, 其中 σ^2 已知, 样本为 X_1, \dots, X_n . 考虑检验水平为 α 的检验问题 ($\mu_2 < \mu_1$),

$$H_0: \mu = \mu_1 \text{ vs. } H_1: \mu = \mu_2$$

解：似然比检验的拒绝域为： $W = \{\vec{x} : \frac{L(\vec{x}; \mu_2)}{L(\vec{x}; \mu_1)} > \lambda\}$. 令似然比

$$\frac{L(X_1, \dots, X_n; \mu_2)}{L(X_1, \dots, X_n; \mu_1)} = \prod_{i=1}^n \frac{f(X_i; \mu_2, \sigma^2)}{f(X_i; \mu_1, \sigma^2)} = e^{\frac{n(\mu_2 - \mu_1)(2\bar{X} - \mu_1 - \mu_2)}{2\sigma^2}}$$

等价于找 C , 使得 $P_{\mu_1}(\bar{X} < C) = \alpha$, $C = \mu_1 + u_\alpha \sigma / \sqrt{n}$, 所以似然比检验的拒绝域为 $W = \{\vec{x} : \bar{x} < \mu_1 + u_\alpha \sigma / \sqrt{n}\}$

单调似然比(Monotone Likelihood Ratio)

定义 3.5. 设样本 X_1, \dots, X_n 的似然函数为 $L(x_{1:n}; \theta)$, 其中 $\theta \in \Theta \subset \mathbb{R}$. 若似然函数的支撑集与 θ 无关, 且存在统计量 $T(X_{1:n}) \in \mathbb{R}$ 使得对任意给定的 $\theta_1 < \theta_2$, 似然比

$$R(\theta_1, \theta_2) := \frac{L(x_{1:n}; \theta_2)}{L(x_{1:n}; \theta_1)} = h(T(x_{1:n}))$$

为关于 T 的单调不减函数, 则称该似然函数具有单调似然比 (关于统计量 T).

定理 3.9. 假设似然函数 $L(x_{1:n}; \theta)$ 具有单调似然比 (关于统计量 T). 令 $W = \{x_{1:n} | T(x_{1:n}) > C\}$, 其中 C 满足 $\rho_W(\theta_0) = \mathbb{P}_{\theta_0}(T(X_{1:n}) > C) = \alpha$. 则 W 是下列检验的显著性水平为 α 的 UMP 拒绝域:

(I) $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1 > \theta_0$

(II) $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$

(III) $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$.

单调似然比(Monotone Likelihood Ratio)

定理 3.10. 假设似然函数 $L(x_{1:n}; \theta)$ 具有单调似然比 (关于统计量 T). 令 $W = \{x_{1:n} | T(x_{1:n}) < C\}$, 其中 C 满足 $\rho_W(\theta_0) = \mathbb{P}_{\theta_0}(T(X_{1:n}) < C) = \alpha$. 则 W 是下列检验的显著性水平为 α 的 UMP 拒绝域:

(I) $H_0 : \theta = \theta_0$ vs. $H_1 : \theta < \theta_0$

(II) $H_0 : \theta = \theta_0$ vs. $H_1 : \theta < \theta_0$

(III) $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$.

此外, 对于双边检验问题

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0,$$

我们可以考虑拒绝域 $W = \{x_{1:n} | T(x_{1:n}) < C_1 \text{ 或 } T(x_{1:n}) > C_2\}$, 其中 $C_1 \leq C_2$ 满足 $\rho_W(\theta_0) = \mathbb{P}_{\theta_0}(T(X_{1:n}) < C_1) + \mathbb{P}_{\theta_0}(T(X_{1:n}) > C_2) = \alpha$. 满足该条件的 C_1 和 C_2 有无穷多种, 为简便起见, 采用平分法, 即 $\mathbb{P}_{\theta_0}(T(X_{1:n}) < C_1) = \mathbb{P}_{\theta_0}(T(X_{1:n}) > C_2) = \alpha/2$. 这样 C_1 和 C_2 分别对应统计量 T 的 $\alpha/2$ 和 $1 - \alpha/2$ 分位数 (在 $\theta = \theta_0$ 下计算). 不过值得注意的是, 这样得到的拒绝域往往不是 UMP 的.

例子

例 3.8. 假设 X_1, \dots, X_n 是来自 $Exp(\lambda)$ 的简单样本. 求下列检验问题的显著性水平为 α 的 UMP 拒绝域.

(I) $H_0 : \lambda \leq \lambda_0$ vs. $H_1 : \lambda > \lambda_0$;

(II) $H_0 : \lambda \geq \lambda_0$ vs. $H_1 : \lambda < \lambda_0$;

(III) $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$.

解. 对任意的 $\lambda_2 > \lambda_1 > 0$, 似然比

$$\frac{L(x_{1:n}; \lambda_2)}{L(x_{1:n}; \lambda_1)} = \prod_{i=1}^n \frac{\lambda_2 \exp(-\lambda_2 x_i)}{\lambda_1 \exp(-\lambda_1 x_i)} = (\lambda_2 / \lambda_1)^n \exp \left[-(\lambda_2 - \lambda_1) \sum_{i=1}^n x_i \right]$$

广义似然比检验

设总体的密度函数为 $f(x; \theta)$, $\theta \in \Theta$, 研究检验问题

$$H_0: \theta \in \Theta_0 \text{ vs. } H_1: \theta \notin \Theta_0$$

设似然函数(likelihood function)为 $L(x_{1:n}; \theta) = \prod_{i=1}^n f(x_i; \theta)$, 定义广义似然比为:

$$\text{GLR} := \frac{\sup_{\theta \in \Theta} L(x_{1:n}; \theta)}{\sup_{\theta \in \Theta_0} L(x_{1:n}; \theta)} = \frac{L(x_{1:n}; \hat{\theta}_{MLE})}{L(x_{1:n}; \hat{\theta}_{MLE}^0)}$$

广义似然比拒绝域为:

$$W = \{x_{1:n} | \text{GLR} > \lambda_0\}$$

其中 λ_0 满足 $\sup_{\theta \in \Theta_0} P_{\theta}(X_{1:n} \in W) = \alpha$, α 为显著性水平。

注意: 这样得到的拒绝域不一定是UMP的!

3. 正态总体的检验

正态总体 $N(\mu, \sigma^2)$ 总体均值和方差的假设检验问题, 考虑下面三种情况:

$$H_0: \theta = \theta_0 \text{ vs. } H_1: \theta \neq \theta_0$$

$$H_0: \theta \leq \theta_0 \text{ vs. } H_1: \theta > \theta_0$$

$$H_0: \theta \geq \theta_0 \text{ vs. } H_1: \theta < \theta_0$$

其中 $\theta = \mu$ 或者 σ^2 , 另外一个参数已知(单参数)或者未知(多参数)。

求解步骤

Step 1. 选择检验方法, 得到检验统计量 $T(X_1, \dots, X_n)$

- 单参数时, 用单调似然比方法
- 多参数时, 用广义似然比方法

Step 2. 求 $\theta = \theta_0$ 下, 检验统计量 $T(X_1, \dots, X_n)$ 的分布

Step 3. 写出拒绝域的形式 $W = \{T(X_1, \dots, X_n) \in R\}$, R 是集合满足

$$\sup_{\theta \in \theta_0} \rho_W(\theta) = P_{\theta_0}(T \in R) = \alpha$$

控制犯第一类错误的概率

Step 4. 代入数据作出判断

注意: 犯第一类错误的概率虽然得到控制, 但并不意味一致最大功效, 很多问题无法得到UMP检验.

Two-sided tests for normal mean

总体 $X_i \sim N(\mu, \sigma^2)$, 其中 σ^2 已知, 即 $\theta = \mu$

似然函数: $L(\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$

单调似然比: 对任意的 $\mu_2 > \mu_1$,

$$\begin{aligned}\frac{L(\mu_2)}{L(\mu_1)} &= \exp\left(\frac{\sum_{i=1}^n (x_i - \mu_1)^2 - \sum_{i=1}^n (x_i - \mu_2)^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{n(\mu_2 - \mu_1)}{2\sigma^2} [2\bar{x} - \mu_1 - \mu_2]\right) \\ &= h(\bar{x}) = \tilde{h}(\mathbf{u})\end{aligned}$$

其中, $\mathbf{u} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

One-sided tests for normal mean II

在方差已知的情况下考虑期望的检验问题, 检验统计量为

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \quad (\text{当}\mu = \mu_0\text{时})$$

该检验统计的直观含义: 衡量 μ 与 μ_0 的偏差

这种检验方式称为u检验(或z检验)

表 3.4: 期望的假设检验 (方差已知)

H_0	$\mu = \mu_0$	$\mu \leq \mu_0$	$\mu \geq \mu_0$
H_1	$\mu \neq \mu_0$	$\mu > \mu_0$	$\mu < \mu_0$
拒绝域	$ u > u_{1-\alpha/2}$	$u > u_{1-\alpha}$	$u < u_\alpha$

Tests for normal mean

期望的假设检验（方差未知）：在方差未知的情况下考虑期望的检验问题，相应的检验统计量为

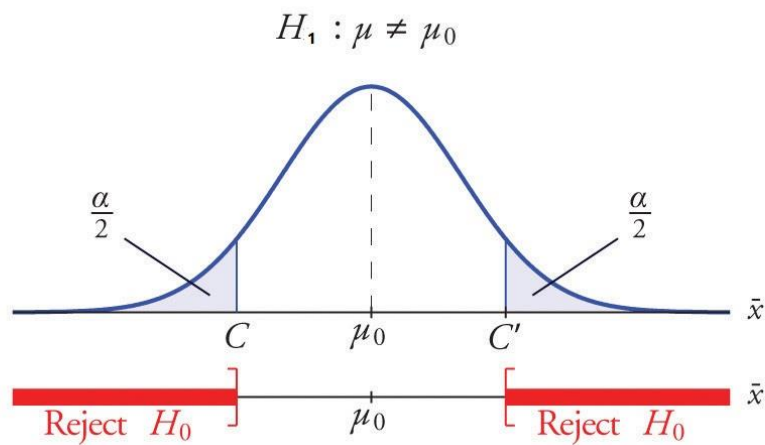
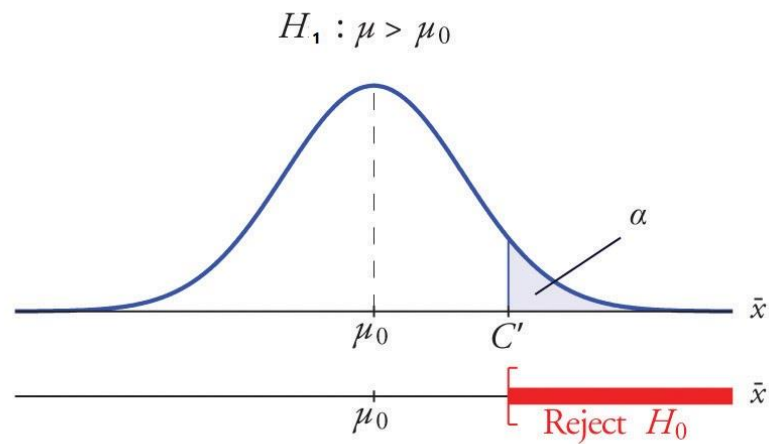
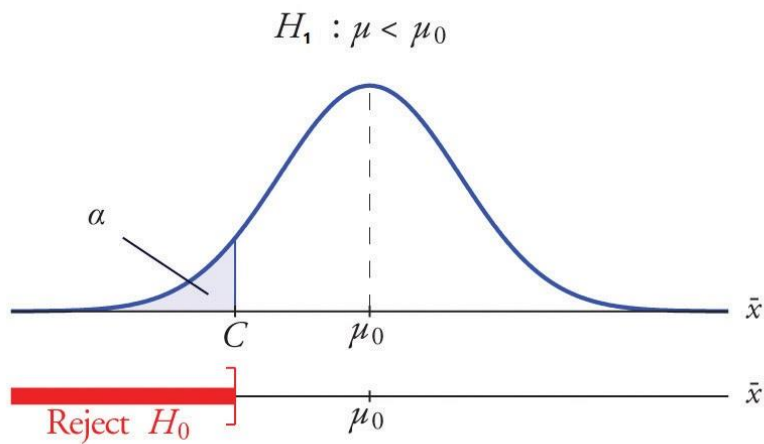
$$T = \frac{\bar{X} - \mu_0}{S_n / \sqrt{n-1}} = \frac{\bar{X} - \mu_0}{S_n^* / \sqrt{n}}.$$

当 $\mu = \mu_0$ 时, $T \sim t(n-1)$. 该检验称为**t检验**。拒绝域 W 形式如下：

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

该检验统计的直观含义：**衡量 μ 与 μ_0 的偏差**

期望的检验



方差的检验

方差的假设检验（期望已知）：检验统计量为

$$V_1 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu)^2.$$

衡量 σ^2 与 σ_0^2 的偏差

当 $\sigma^2 = \sigma_0^2$ 时, $V_1 \sim \chi^2(n)$ 。该检验称为**卡方检验**。拒绝域 W 形式如下：

H_1	$\sigma^2 \neq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\sigma^2 < \sigma_0^2$
拒绝域	$v_1 > \chi_{1-\alpha/2}^2(n)$ 或 $v_1 < \chi_{\alpha/2}^2(n)$	$v_1 > \chi_{1-\alpha}^2(n)$	$v_1 < \chi_{\alpha}^2(n)$

方差的假设检验（期望未知）：检验统计量为

$$V_2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 = nS_n^2/\sigma_0^2.$$

衡量 σ^2 与 σ_0^2 的偏差

当 $\sigma^2 = \sigma_0^2$ 时, $V_2 \sim \chi^2(n-1)$ 。该检验同样为**卡方检验**。拒绝域 W 形式如下：

H_1	$\sigma^2 \neq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\sigma^2 < \sigma_0^2$
拒绝域	$v_2 > \chi_{1-\alpha/2}^2(n-1)$ 或 $v_2 < \chi_{\alpha/2}^2(n-1)$	$v_2 > \chi_{1-\alpha}^2(n-1)$	$v_2 < \chi_{\alpha}^2(n-1)$

血液酒精浓度测试

例 3.9 (血液酒精浓度测试). 下面是一台已使用三年的仪器测出某人血液酒精浓度的 30 个数据 (百分比). 已知精准的机器给出读数为 12.6%. 请根据这些数据检验这台仪器是否精准, 是否需要校准.

表 3.8: 血液酒精浓度测试数据

12.3	12.7	13.6	12.7	12.9	12.6
12.6	13.1	12.6	13.1	12.7	12.5
13.2	12.8	12.4	12.6	12.4	12.4
13.1	12.9	13.3	12.6	12.6	12.7
13.1	12.4	12.4	13.1	12.4	12.9

如果这台老仪器的方差 $\sigma^2 = 0.4$. 在显著性水平 $\alpha = 0.05$ 下, 你是否建议对该仪器进行校准? 如果显著性水平减小到 $\alpha = 0.01$, 你的结论会不会发生改变?

求解步骤

解. 假设这台仪器读数服从正态分布 $N(\mu, \sigma^2)$. 题中问题可描述成

$$H_0 : \mu = 12.6 \text{ vs. } H_1 : \mu \neq 12.6.$$

由数据知, $\bar{x} = 12.757$. 拒绝域为

$$W = \{|\bar{x} - 12.6| > u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\}.$$

如果 $\alpha = 0.05$, $W = \{|\bar{x} - 12.6| > 0.143\}$. 此时, 样本落在拒绝域里面, 故建议机器进行校准。如果 α 减小到 0.01, 拒绝域为 $W = \{|\bar{x} - 12.6| > 0.188\}$. 此时, 样本落在拒绝域外, 故不建议机器进行校准。

t检验（方差未知）：R语言求解

```
t.test(x, alternative = c("two.sided", "less", "greater"), mu = 0, conf.level = 0.95, ...)
```

```
x = c(12.3, 12.7, 13.6, 12.7, 12.9, 12.6,  
      12.6, 13.1, 12.6, 13.1, 12.7, 12.5,  
      13.2, 12.8, 12.4, 12.6, 12.4, 12.4,  
      13.1, 12.9, 13.3, 12.6, 12.6, 12.7,  
      13.1, 12.4, 12.4, 13.1, 12.4, 12.9)
```

```
t.test(x, mu=12.6)
```

结果解读

```
##  
## One Sample t-test  
##  
## data: x  
## t = 2.6444, df = 29, p-value = 0.01307  
## alternative hypothesis: true mean is not equal to 12.6  
## 95 percent confidence interval:  
## 12.63550 12.87784  
## sample estimates:  
## mean of x  
## 12.75667
```

因为 $|t| = 2.6444 > t_{0.975}(29) = 2.04523$, 所以在显著性水平 $\alpha = 0.05$ 下拒绝 H_0 。但 $|t| = 2.6444 < t_{0.995}(29) = 2.756386$, 所以在显著性水平 $\alpha = 0.01$ 下接受 H_0 。结论与之前一致。

鸢尾花例子

1. 假设有个植物学家跟你说，通过基因组分析表明，山鸢尾(Setosa)花萼长度的均值是4.5cm，他这个论断是否可信？
2. 另一位植物学家说山鸢尾(Setosa)花萼长度的均值是一个不小于4.5cm的数，但具体是多少就不清楚了，那么他这个论断是否又可信？

R语言求解结果

```
x = iris[iris$Species=="setosa",1]
```

```
t.test(x,mu=4.5)
```

```
##
## One Sample t-test
##
## data:  x
## t = 10.151, df = 49, p-value = 1.223e-13
## alternative hypothesis: true mean is not equal to 4.5
## 95 percent confidence interval:
##  4.905824 5.106176
## sample estimates:
## mean of x
##      5.006
```

拒绝H0

```
t.test(x,mu=4.5, alternative = "less")
```

```
##
## One Sample t-test
##
## data:  x
## t = 10.151, df = 49, p-value = 1
## alternative hypothesis: true mean is less than 4.5
## 95 percent confidence interval:
##      -Inf 5.089575
## sample estimates:
## mean of x
##      5.006
```

接受H0

两个独立正态总体的检验

有时候，我们需要比较两个总体的差异性。例如，男生群体平均身高是否大于女生群体平均身高？山鸢尾和杂色鸢尾两种花的花萼有没有显著差异？大部分问题归结为比较两个总体均值或者方差的差异性。

前提条件： 设总体 $X \sim N(\mu_1, \sigma_1^2)$, 另有与 X 独立的总体 $Y \sim N(\mu_2, \sigma_2^2)$. 两个总体的样本分别为 $X_1, \dots, X_m; Y_1, \dots, Y_n$, 修正样本方差分别为 S_X^{*2}, S_Y^{*2} .

1. 考虑检验问题：

$$H_0 : \mu_1 - \mu_2 = \delta \text{ vs. } H_1 : \mu_1 - \mu_2 \neq \delta.$$

大部分应用场景取 $\delta = 0$.

• 如果 σ_1^2, σ_2^2 已知, 选择**U检验统计量**：

$$U = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \stackrel{H_0}{\sim} N(0, 1).$$

若 $|U| > u_{1-\alpha/2}$ 拒绝 H_0 , 否则接受 H_0 .

两个独立正态总体的检验

- 如果 σ_1^2, σ_2^2 未知, 已知 $\sigma_1^2 = \sigma_2^2$, 选择**t检验统计量**:

$$T = \frac{\bar{X} - \bar{Y} - \delta}{S_w \sqrt{1/m + 1/n}} \stackrel{H_0}{\sim} t(m + n - 2),$$

其中 $S_w^2 = [(m-1)S_X^{*2} + (n-2)S_Y^{*2}]/(m+n-2)$ 为合并的样本方差。

若 $|T| > t_{1-\alpha/2}(n+m-2)$ 拒绝 H_0 , 否则接受 H_0 .

- 如果 σ_1^2, σ_2^2 未知, 但 $\sigma_1^2 \neq \sigma_2^2$, 选择**检验统计量**:

$$T = \frac{(\bar{X} - \bar{Y}) - \delta}{\sqrt{S_X^{*2}/m + S_Y^{*2}/n}}.$$

在 $\mu_1 - \mu_2 = \delta$ 下, T 近似服从自由度为 k 的 t 分布, 其中 k 为接近 k^* 的整数,

$$k^* = \frac{(S_X^{*2}/m + S_Y^{*2}/n)^2}{(S_X^{*2}/m)^2/(m-1) + (S_Y^{*2}/n)^2/(n-1)}$$

这就是著名的**Behrens-Fisher问题**, 该检验为**Welch's t-test**。

两个独立正态总体的检验

配对检验：假设两个总体的样本量都一样，即 $m = n$ 。此时，令

$$Z = X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2),$$

Z 的样本为 $Z_i = X_i - Y_i, i = 1, \dots, n$. 原问题可以转化成关于 Z 的均值的检验，故可构造**t检验统计量**

$$T = \frac{\bar{Z} - \delta}{S_z^*/\sqrt{n}} \sim t(n-1),$$

其中 S_z^{*2} 为 Z_i 的修正样本方差。这种检验方法称**配对检验(paired test)**，优点是对两个总体的方差没有要求，甚至不要求两个总体独立。但只适用于 $m = n$ 的情况。如果 $m \neq n$ ，这种配对检验就不太适合，尤其是两个样本量相差很大时，做配对检验就不得不舍弃大量的样本，造成一定的信息丢失。

两个独立正态总体的检验

2. 考虑方差比值的检验问题

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

- 如果 μ_1, μ_2 已知, 选择**F检验统计量**:

$$F_1 = \frac{\frac{1}{m} \sum_{i=1}^m (X_i - \mu_1)^2}{\frac{1}{n} \sum_{i=1}^n (Y_i - \mu_2)^2} \stackrel{H_0}{\sim} F(m, n).$$

若 $F_1 > F_{1-\alpha/2}(m, n)$ 或者 $F_1 < F_{\alpha/2}(m, n)$ 拒绝 H_0 , 否则接受 H_0 .

- 如果 μ_1, μ_2 未知, 选择**F检验统计量**:

$$F_2 = \frac{\frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2}{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{S_X^{*2}}{S_Y^{*2}} \stackrel{H_0}{\sim} F(m-1, n-1) \quad .$$

若 $F_2 > F_{1-\alpha/2}(m-1, n-1)$ 或者 $F_2 < F_{\alpha/2}(m-1, n-1)$ 拒绝 H_0 , 否则接受 H_0 .

山鸢尾和杂色鸢尾花差异性比较

均值的假设检验R命令

```
t.test(x, y = NULL, alternative = c("two.sided", "less", "greater"), mu = 0,  
paired = FALSE, var.equal = FALSE, conf.level = 0.95, ...)
```

方差的假设检验R命令

```
var.test(x, y, ratio = 1, alternative = c("two.sided", "less", "greater"),  
conf.level = 0.95, ...)
```

1. 假设山鸢尾和杂色鸢尾两种花的花萼长度方差相等时，花萼长度均值的双边假设检验

```
x = iris[iris$Species=="setosa",1]
y = iris[iris$Species=="versicolor",1]
t.test(x,y,var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data:  x and y
## t = -10.521, df = 98, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.1054165 -0.7545835
## sample estimates:
## mean of x mean of y
##      5.006      5.936
```

因为 $|t| = 10.521 > t_{0.975}(98) = 1.984467$ ，所以在显著性水平 $\alpha = 0.05$ 下拒绝 H_0 ，即认为两种花的花萼长度均值相等不显著。

2. 方差不相等时双边假设检验，即Welch的t检验。

```
t.test(x,y)
```

```
##  
## Welch Two Sample t-test  
##  
## data:  x and y  
## t = -10.521, df = 86.538, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
##  -1.1057074 -0.7542926  
## sample estimates:  
## mean of x mean of y  
##      5.006      5.936
```

因为 $|t| = 10.521 > t_{0.975}(87) = 1.987608$ ，在显著性水平 $\alpha = 0.05$ 下同样拒绝 H_0 ，即认为两种花的花萼长度均值相等不显著。

3. 方差不相等时单边假设检验

```
t.test(x,y,alternative = "less")
```

```
##  
##  Welch Two Sample t-test  
##  
## data:  x and y  
## t = -10.521, df = 86.538, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
##      -Inf -0.7830302  
## sample estimates:  
## mean of x mean of y  
##      5.006      5.936
```

因为 $t = -10.521 < t_{0.05}(87) = -1.662557$, 在显著性水平 $\alpha = 0.05$ 下拒绝 H_0 , 即认为山鸢尾花萼平均长度小于杂色鸢尾花萼平均长度。

4. 山鸢尾和杂色鸢尾两种花的花萼长度方差比的检验

```
var.test(x,y)
```

```
##  
##  F test to compare two variances  
##  
## data:  x and y  
## F = 0.46634, num df = 49, denom df = 49, p-value = 0.008657  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
##  0.2646385 0.8217841  
## sample estimates:  
## ratio of variances  
##           0.4663429
```

因为 $F = 0.46634 < F_{0.025}(49, 49) = 0.5674762$, 在显著性水平 $\alpha = 0.05$ 下拒绝 H_0 , 即认为两种花的花萼长度方差相等是不显著的。 所以一开始认为两者方差相等对均值进行检验是不合适的。

5. 通过配对检验山鸢尾和杂色鸢尾两种花的花萼长度平均水平的差异

```
t.test(x,y,paired = TRUE)
```

```
##  
## Paired t-test  
##  
## data: x and y  
## t = -10.146, df = 49, p-value = 1.242e-13  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.114203 -0.745797  
## sample estimates:  
## mean of the differences  
## -0.93
```

因为 $|t| = 10.146 > t_{0.975}(49) = .009575$ ，所以在显著性水平 $\alpha = 0.05$ 下拒绝 H_0 ，即认为两种花的花萼长度均值相等不显著（与Welch的t检验结果一致）。

Confidence intervals vs HT

假设 θ 的 $100(1 - \alpha)\%$ 置信区间(confidence interval, CI)为 $[L(X_1, \dots, X_n), U(X_1, \dots, X_n)]$. 这表明

$$P_{\theta}(\theta \in [L, U]) = 1 - \alpha, \forall \theta \in \Theta.$$

考虑假设检验:

$$H_0: \theta = \theta_0 \text{ vs. } H_1: \theta \neq \theta_0$$

检验法则: 如果 $\theta_0 \notin [L, U]$, 拒绝原假设; 否则接受原假设。于是得到一个拒绝域: $W = \{\vec{x}: \theta_0 \notin [L(\vec{x}), U(\vec{x})]\}$, 显著性水平为

$$P_{\theta_0}(\theta_0 \notin [L, U]) = \alpha$$

但这样得到的拒绝域不一定是UMP!

Confidence intervals vs HT

反过来，假如我们有以下检验的一个拒绝域 $W(\theta_0)$

$$H_0: \theta = \theta_0 \text{ vs. } H_1: \theta \neq \theta_0,$$

其中 $P_{\theta_0}((X_1, \dots, X_n) \in W(\theta_0)) = \alpha, \forall \theta_0 \in \Theta$. 可以得到一个置信集
(**confidence set**):

$$S(X_1, \dots, X_n) = \{\theta: (X_1, \dots, X_n) \notin W(\theta)\}$$

$$P_{\theta}(\theta \in S) = P_{\theta}((X_1, \dots, X_n) \notin W(\theta)) = 1 - \alpha, \forall \theta \in \Theta$$

- 该置信集是由所有“接受”的 θ 的值组成的
- 一般情况下，该置信集为区间形式

注：这种关系称为假设检验与置信区间的**对偶性(duality)**

即由拒绝域可以诱导出置信区间，由置信区间可以诱导出拒绝域。

Example

考虑总体 $X \sim N(\mu, \sigma^2)$ ，其中 μ 未知，方差 σ^2 已知。则有

$$P\left(\bar{X} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

于是得到一个置信区间： $[\bar{X} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}]$

由此可以构造拒绝域：

$$W = \{\mu_0 \notin [\bar{x} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}]\}$$

$$W = \{\vec{x} : |\bar{x} - \mu_0| > u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\}$$

此拒绝域和我们之前得到的一样的。

p-values

无糖型 骨松宝颗粒说明书

GUSONGBAO KELI SHUOMINGSHU

请仔细阅读使用说明书并在医师指导下使用

- 【药品名称】 骨松宝颗粒
通用名称: Gusongbao Keli
汉语拼音: 淫羊藿、续断、赤芍、川芎、知母、莪术、三棱、地黄、牡蛎(煅)。
- 【性状】 本品为淡棕色至棕褐色的颗粒; 气微香, 味微苦。
- 【功能主治】 补肾活血, 强筋壮骨。用于骨痿(骨质疏松)引起的骨折、骨痛、骨关节炎, 以及预防更年期骨质疏松。
- 【规格】 每袋5克。
- 【用法用量】 口服, 一次1袋; 治疗骨折及骨3次; 预防骨质疏松, 一日2次, 30。
- 【不良反应】 尚不明确
- 【禁忌】 尚不明确
- 【注意事项】 尚不明确
- 【药理毒理】 药理作用:

本品具有明显的抗炎止痛、提
改善“阳虚”症等作用, 具有
谢, 提高骨密度, 改善骨质
毒理研究:
骨松宝颗粒一日内给一级KM小鼠
最大耐受量为108g/kg, 按体重
用药量的216倍, 为无毒性中毒危
松宝颗粒以10g·kg⁻¹、20g·kg⁻¹
的剂量(相当于临床用药量的20倍、40倍、80
倍)连续给一级SD大鼠灌胃给药26周, 整个给
药期间无动物死亡, 各组动物活动、行为、
大小便正常; 给药组动物体重增长、外周血
象、肝肾功能、内脏系数与对照组比较, 无
明显差异(P>0.05); 活杀动物的解剖学和
病理组织学检查均未见特殊病理变化, 停药
两周的各项检查亦未见异常。

- 【贮藏】 密封。
- 【包装】 包装材料: 铝箔复合膜。6袋/盒。
- 【有效期】 24个月
- 【执行标准】 部颁标准中药成方制剂第十七册、WS3-B-3292-98
- 【批准文号】 国药准字Z52020005
- 【生产企业】

企业名称: 贵州富华药业有限责任公司
生产地址: 贵州省龙里县谷脚镇岩后村医药工业城
邮政编码: 551206
电话号码: 0854-5670608
传真号码: 0854-5670041
注册地址: 贵州省龙里县谷脚镇岩后村医药工业城
网 址: www.fuhuayy.com

如有问题可与生产企业直接联系

倍)连续给一级SD大鼠灌胃给药26周, 整个给
药期间无动物死亡, 各组动物活动、行为、
大小便正常; 给药组动物体重增长、外周血
象、肝肾功能、内脏系数与对照组比较, 无
明显差异(P>0.05); 活杀动物的解剖学和
病理组织学检查均未见特殊病理变化, 停药
两周的各项检查亦未见异常。

p-values

假设拒绝域具备如下形式

$$W = \{\vec{x} : T(\vec{x}) > \lambda\}$$

- T 为检验统计量
- λ_α 满足 $\sup_{\theta \in \Theta_0} P_\theta(T(\vec{X}) > \lambda_\alpha) = \alpha$

由此可得：

- 对于固定的样本，显著性水平 α 越大， λ_α 越小，这样越容易拒绝原假设
- 对于固定的样本，是否存在一个临界值 p ，使得当 $p < \alpha$ 时拒绝原假设，当 $p \geq \alpha$ 时接受原假设？这个临界值称为 p 值(p-value)

$$p = p(\vec{x}) = \sup_{\theta \in \Theta_0} P_\theta(T(X_1, \dots, X_n) > T(x_1, \dots, x_n))$$

如果给定 α , λ_α 存在且唯一，可证明： $T(\vec{x}) > \lambda_\alpha$ 当且仅当 $p(\vec{x}) < \alpha$

p-values for simple null

如果原假设是简单的, 即 $H_0: \theta = \theta_0$, 则

$$p = P_{\theta_0}(T(X_1, \dots, X_n) > T(x_1, \dots, x_n))$$

例: 总体 $X \sim N(\mu, \sigma^2)$, 其中 μ 未知, 方差 σ^2 已知, 考虑检验

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

拒绝域为 $W = \{\vec{x}: \sqrt{n}|\bar{x} - \mu_0|/\sigma > u_{1-\alpha/2}\}$, 其中检验统计量为

$T = \sqrt{n}|\bar{X} - \mu_0|/\sigma$. 故 p 值为

$$p = P_{\mu_0}(T(\vec{X}) > T(\vec{x})) = 2 - 2\Phi(T(\vec{x})) = 2 - 2\Phi\left(\frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right)$$

For blood alcohol determinations example,

$$p = 2 - 2\Phi\left(\frac{|12.757 - 12.6|}{0.4/\sqrt{30}}\right) = \mathbf{0.032}.$$

Comments

- p值可以看作样本与原假设**相容程度的度量**。p值越大相容度越高；反之，p值越小相容度越低。当p值小于 α 时认为两者不相容，拒绝原假设
- 做检验时不需要事先确定显著性水平 α （它的具有一定的主观性），如果p值非常小，则毫不犹豫地拒绝原假设；同样地，如果p值比较大，则接受原假设，这样就不用争论 $\alpha = 0.1, 0.05$ 或者其他。
- p值提供更多的信息，可以用于**保护隐私数据**
- 统计软件提供的是p值

Binomial tests

设 X 服从两点分布 $B(1, p)$, 下面考虑以下三种常见的假设检验

- $H_0: p \leq p_0$ vs. $H_1: p > p_0$
- $H_0: p \geq p_0$ vs. $H_1: p < p_0$
- $H_0: p = p_0$ vs. $H_1: p \neq p_0$

对于该总体, 我们选 $S = \sum_{i=1}^n X_i \sim B(n, p)$ 为检验统计量。相应的拒绝域形式为

- $W = \{s \geq c\}$
- $W = \{s \leq c\}$
- $W = \{s \geq c_2\} \cup \{s \leq c_1\}$

注意到 S 为离散型随机变量, 所以满足 $\sup_{p \in \theta_0} P_p(\vec{X} \in W) = \alpha$ 的分界点不一定存在。因此, 我们考虑 $\sup_{p \in \theta_0} P_p(\vec{X} \in W) \leq \alpha$ 下分界点的选取。

HT for ratio: one-sided I

考虑 $H_0: p \leq p_0$ vs. $H_1: p > p_0$, 临界值 c 为满足下式最小的整数

$$\sup_{p \leq p_0} P_p(S \geq c) \leq \alpha$$

引理 3.1 设 $F_\beta(x; a, b)$ 为 $\text{Beta}(a, b)$ 分布的累积分布函数, 则

$$P_p(S \geq c) = \sum_{i=c}^n C_n^i p^i (1-p)^{n-i} = F_\beta(p; c, n-c+1),$$

其中 $F_\beta(p; c, n-c+1)$ 表示 $\text{Beta}(c, n-c+1)$ 分布的CDF在 p 点处的取值。

$$P_{p_0}(S \geq c) = \sum_{i=c}^n C_n^i p_0^i (1-p_0)^{n-i} \leq \alpha$$

计算 c 比较复杂, 为了避免此, 我们将拒绝域 $\{s \geq c\}$ 等价转化为

$$W = \left\{ \sum_{i=s}^n C_n^i p_0^i (1-p_0)^{n-i} \leq \alpha \right\}$$

HT for ratio: one-sided I

更进一步，假设 $p_\alpha(s)$ 为方程 $\sum_{i=s}^n C_n^i p^i (1-p)^{n-i} = \alpha$ 的根，则拒绝域等价转化为

$$W = \{p_0 \leq p_\alpha(s)\}$$

其中 $p_\alpha(s) = \text{Beta}_\alpha(s, n-s+1)$, 或者可以表示成

$$p_\alpha(s) = \left(1 + \frac{n-s+1}{s} F_{1-\alpha}(2(n-s+1), 2s)\right)^{-1}$$

Example

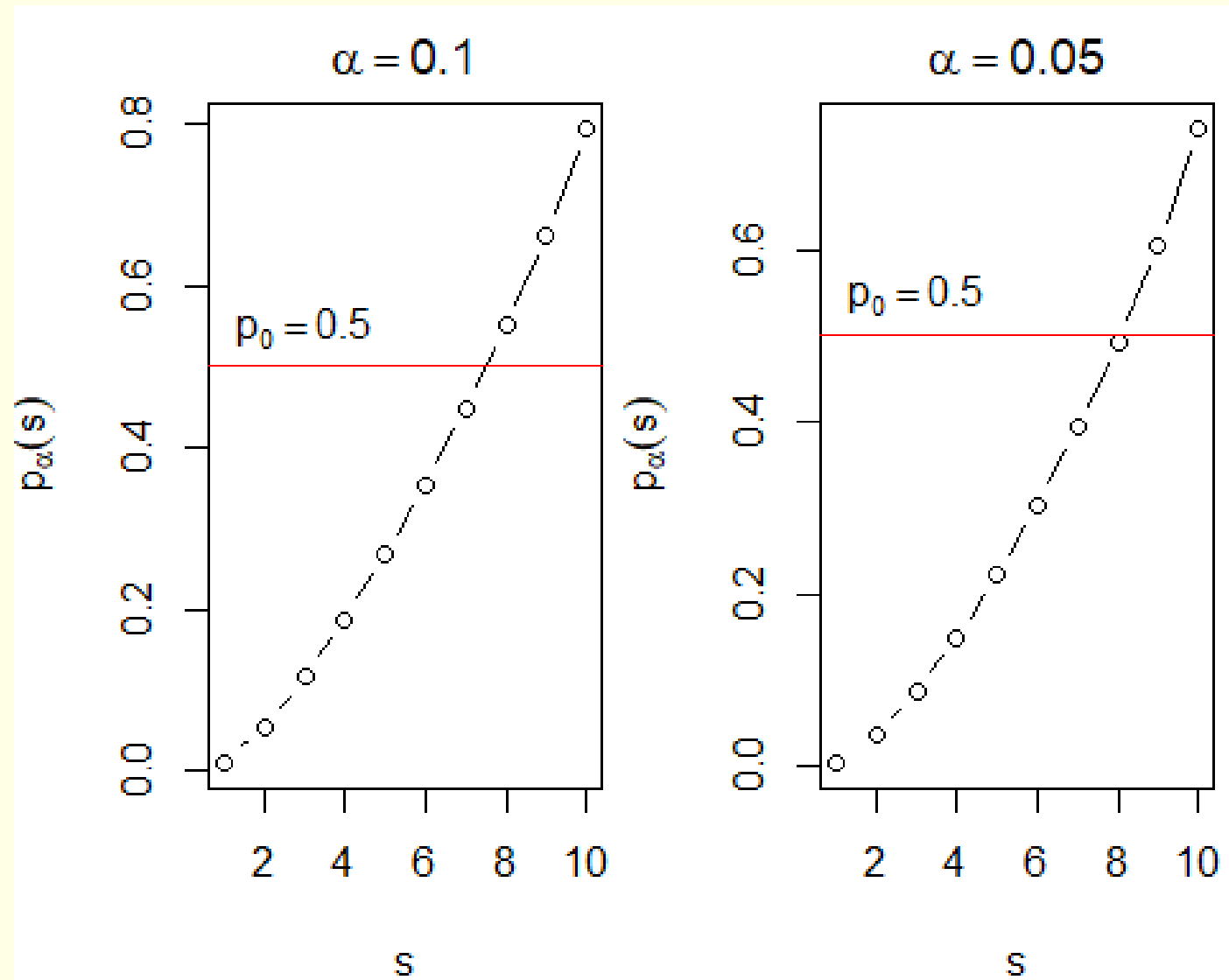
考虑女士品茶问题，设该女士鉴别的成功率为 p . 设 X_i 表示第 i 次鉴别结果，即 $X_i = 1$ 表示成功， $X_i = 0$ 表示失败。如果 $p > p_0$ 我们认为该女士具备这种辨别能力，其中 $p_0 \geq 1/2$ 为给定的数。故考虑检验

$$H_0: p \leq p_0 \text{ vs. } H_1: p > p_0.$$

二项分布检验的R代码

```
binom.test(x, n, p = 0.5, alternative = c("two.sided", "less", "greater"),  
conf.level = 0.95)
```

The results for $n=10$



binom.test

```
##  
##  Exact binomial test  
##  
## data:  8 and 10  
## number of successes = 8, number of trials =  
10, p-value = 0.05469  
## alternative hypothesis: true probability of  
success is greater than 0.5  
## 95 percent confidence interval:  
##  0.4930987 1.0000000  
## sample estimates:  
## probability of success  
##                                0.8
```

HT for ratio: one-sided II

考虑单边假设 $H_0: p \geq p_0$ vs. $H_1: p < p_0$, 临界值 c 为满足下式最大的整数

$$\sup_{p \geq p_0} P_p(S \leq c) \leq \alpha$$

由于 $P_p(S \leq c) = \sum_{i=0}^c C_n^i p^i (1-p)^{n-i}$ 关于 p 单调递减, 所以只需考虑

$$P_{p_0}(S \leq c) = \sum_{i=0}^c C_n^i p_0^i (1-p_0)^{n-i} \leq \alpha$$

计算 c 比较复杂, 为了避免此, 我们将拒绝域 $\{s \geq c\}$ 等价转化为

$$W = \left\{ \sum_{i=s+1}^n C_n^i p_0^i (1-p_0)^{n-i} \geq 1 - \alpha \right\}$$

HT for ratio: one-sided II

更进一步, 假设 $\tilde{p}_\alpha(s)$ 为方程 $\sum_{i=s+1}^n C_n^i p^i (1-p)^{n-i} = 1 - \alpha$ 的根, 则拒绝域等价转化为

$$W = \{p_0 \geq \tilde{p}_\alpha(s)\}$$

其中 $\tilde{p}_\alpha(s) = \text{Beta}_{1-\alpha}(s+1, n-s)$, 或者可以表示成

$$\tilde{p}_\alpha(s) = \left(1 + \frac{n-s}{(s+1)F_{1-\alpha}(2(s+1), 2(n-s))}\right)^{-1}$$

详细的转化见课本P105引理4.2.

HT for ratio: two-sided

考虑双边假设

$$H_0: p = p_0 \text{ vs. } H_1: p \neq p_0$$

拒绝域为 $\{s \leq c_1\} \cup \{s \geq c_2\}$, 其中临界值 c_1 为满足 $P_{p_0}(S \leq c_1) = \alpha/2$ 最大的整数, 临界值 c_2 为满足 $P_{p_0}(S \geq c_2) = \alpha/2$ 最小的整数。由前面分析, 该拒绝域等价于

$$\{p_0 \leq p_{\alpha/2}(s)\} \cup \{p_0 \geq \tilde{p}_{\alpha/2}(s)\}$$

Example: Mendel's Data

In one of his famous experiments, Mendel (孟德尔) crossed 556 smooth, yellow male peas (豌豆) with wrinkled, green female peas. The counts that Mendel recorded are

smooth yellow	smooth green	wrinkled yellow	wrinkled green
315	108	102	31

According to now established genetic theory, the relative frequencies of the progeny should be as given below.

$$P(\text{smooth yellow}) = 9/16, P(\text{smooth green}) = 3/16$$

$$P(\text{wrinkled yellow}) = 3/16, P(\text{wrinkled green}) = 1/16$$

Would you conclude that Mendel's experiment is correct at the level of significance $\alpha = 0.05$?

4. Goodness-of-fit tests (拟合优度检验)

考虑离散型分布的假设检验, $X \in \{t_1, \dots, t_m\}$.

$$H_0: P(X = t_i) = p_i, i = 1, \dots, m, \text{ vs. } H_1: P(X = t_i) \neq p_i$$

其中 $\sum_{i=1}^m p_i = 1$.

Pearson卡方检验法检验统计量:

$$V = \sum_{i=1}^m \frac{(v_i - np_i)^2}{np_i}$$

其中 v_i 表示 X_1, \dots, X_n 中包含 t_i 的个数, 即 $v_i = \sum_{j=1}^n 1\{X_j = t_i\}$.

拒绝域 $W = \{V > \lambda\}$. 可以证明在 H_0 下, $V \sim \chi^2(m-1)$, 故 $\lambda = \chi^2_{1-\alpha}(m-1)$

V 里面的项可以理解为相对平方误差。

Example: Mendel's Data

t_i	smooth yellow	smooth green	wrinkled yellow	wrinkled green
v_i	315	108	102	31
p_i	9/16	3/16	3/16	1/16
np_i	312.75	104.25	104.25	34.75

- the χ^2 test statistic is $V = 0.604$, $\chi^2_{1-0.05}(3) = 7.81$, accept the null.
- p-value is 0.90
- R code:

x = c(315, 108, 102, 31)

p = c(9/16, 3/16, 3/16, 1/16)

#注意第二参数要标明是“p=”，不能省略

chisq.test(x,p=p)

chisq.test

```
##  
##  Chi-squared test for given probabilities  
##  
## data:  x  
## X-squared = 0.60432, df = 3, p-value =  
0.8954
```

Goodness-of-fit tests

考虑连续型分布：

$$H_0: F(x) = F_0(x) \text{ vs. } F(x) \neq F_0(x)$$

把整个实轴分成 m 份, $(-\infty, t_1], (t_1, t_2], \dots, (t_{m-2}, t_{m-1}], (t_{m-1}, \infty)$, 分别计算这 m 个区间的概率 $p_i, i = 1, \dots, m$, v_i 表示 X_1, \dots, X_n 落到第 i 个区间的个数, 类似离散的分布的检验。

t_i, m 的选择? 借鉴直方图法的选取

Next?

其他检验：

- 独立性检验
- 正态性检验
- 柯尔莫哥洛夫检验法

Multiple tests

- 如果独立检验同一个假设 k 次，我们可以得到 k 个p值: p_1, \dots, p_k , 可否由这 k 个p值汇总成一个p值来检验该假设？**元分析(meta-analysis)**
- 假如我们有 k 个不同的原假设 $H_{0j}, j = 1, \dots, k$, 这种问题称为**多重假设(multiple tests)**问题。可否利用 k 个不同假设的p值: p_1, \dots, p_k 来进一步控制错误的发生概率？

例子：吃果冻与长青春痘的联系：<https://xkcd.com/882/>

总结

- 假设检验问题与两类错误息息相关
- 但很多问题只能保证第一类错误得到控制（即认为第一类错误重要）
- 警惕犯第二类错误（把你认为正确的结论放在备择假设）
- 熟练运用（单调）似然比检验方法、N-P引理的理解
- 熟练掌握正态总体的检验（单样本、两样本）：U/t/卡方/F检验
- 实际检验问题的解决步骤：
 1. 假设关心的总体服从正态分布，根据题意给出 H_0 和 H_1
 2. 选择检验统计量（U/t/卡方/F检验）
 3. 根据假设的类型确定拒绝域（左侧、右侧、双侧）
 4. 代入数据判断接受或者拒绝 H_0 ，并正面回答题目的问题。