

Chap 7 假设检验

1 基本概念

- **例(女士品茶)** 某女士声称可区分奶茶是先加牛奶还是先加茶. **R.Fisher** 设计实验: 各有 4 杯奶茶随机排成一排, 将这一信息告知女士. 考虑假设 H : 该女士没有区分能力. 当 H 正确的情况下, 4 杯全对的概率为

$$\frac{C_4^4 \cdot C_4^0}{C_8^4} = \frac{1}{70}.$$

下述两种情况之一必发生:

- H 不正确 (i.e. 该女士有区分能力);
- H 正确 (发生了一件概率为 $\frac{1}{70}$ 的事情).

通常选择阈值 $\alpha = 0.05, 0.01, 0.1$ (预先给定的显著性水平). 若女士选对了三杯, 则在 H 正确的前提下, 挑对三杯及以上的概率为

$$\frac{C_4^4 \cdot C_4^0}{C_8^4} + \frac{C_4^3 \cdot C_4^1}{C_8^4} = \frac{17}{70} \approx 0.243.$$

- **注**

- **Fisher** 显著性检验;
- 若认可某组观测(样本), 则用它来证实或证伪某个理论(断言)具有天然的不对等;
- H 可以模型化:

$$P(X = k) = \frac{C_4^k \cdot C_4^{4-k}}{C_8^4}.$$

- 历史注记:

- **Fisher** 显著性检验;
- **Neyman-Pearson** 检验;
- 零假设显著性检验 (**MHST**).

- **定义(统计假设)** 对一个或多个总体的某种断言或猜测.

- 原假设: 被检验的假设 H_0 ;
- 备择假设: 拒绝 H_0 后可供选择的假设 H_1 .
- 若假设可表为参数形式, 那么 $H_0: \theta \in \Theta_0$, $H_1: \theta \in \Theta_1$, 且 $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \cup \Theta_1 = \theta$ 的所有可能取值之集.

- **例** 假设 $X \sim N(\mu, \sigma^2)$, 其中 σ^2 已知.

- $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$. (双侧假设)

- $H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$. (单侧假设)
- $H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0$. (单侧假设)
- $H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$. (双侧假设)

● 注

- 简单假设: 只对应一个总体;
- 复合假设: 对应多个总体;
- 若 σ^2 未知, 则 $H_0 : \mu = \mu_0 \Leftrightarrow H_0 : \mu = \mu_0, \sigma^2$ 任意, 是一个复合假设.

- 定义(假设检验) 依据样本(观测)的决策(拒绝或不拒绝 H_0)过程.
- 定义(检验准则) 做出决策的一个具体法则.
- 定义(拒绝) 在原假设 H_0 为真的前提下, 所观测的样本出现的概率如果是很小的, 意味着样本提供的概率拒绝 H_0 .
- 定义(拒绝域/临界域) 形式上可抽象为

$$R = \{(X_1, \dots, X_n) \mid T(X_1, \dots, X_n) \geq c\}.$$

其中 c 被定义为临界值. 此时检验准则为, 若样本 $(X_1, \dots, X_n) \in R$, 则拒绝假设 H_0 .

- 定义(显著性检验) 对事先给定的 $\alpha \in (0, 1)$, 若 $P_\theta(T(X_1, \dots, X_n) \geq c) \leq \alpha, \forall \theta \in \Theta_0$, 则称这是一个水平为 α 的显著性检验.
- 例 假设 $X \sim N(\mu, \sigma^2)$, 其中 σ^2 已知, 给定检验水平 $\alpha \in (0, 1)$ 和已观测样本 $X_i (1 \leq i \leq n)$. $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$. 对原假设进行检验.
- 解答 拒绝域为双侧拒绝, 当 H_0 为真时, 控制 $P(|\bar{X} - \mu_0| \geq c) \leq \alpha$.

根据 CLT, 注意到

$$\bar{X} - \mu_0 \sim N(0, \frac{\sigma^2}{n}) \Leftrightarrow \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1).$$

$$\text{取 } \frac{c}{\frac{\sigma}{\sqrt{n}}} = Z_{\frac{\alpha}{2}} \Rightarrow c = \frac{\sigma}{\sqrt{n}} \cdot Z_{\frac{\alpha}{2}}.$$

给出在 α 的检验水平下拒绝 H_0 的条件: 若 $|\bar{X} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} \cdot Z_{\frac{\alpha}{2}}$, 则拒绝 H_0 .

- 例 假设 $X \sim N(\mu, \sigma^2)$, 其中 σ^2 已知, 给定检验水平 $\alpha \in (0, 1)$ 和已观测样本 $X_i (1 \leq i \leq n)$. $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$. 对原假设进行检验.
- 解答 拒绝域为单侧拒绝, 当 H_0 为真时, 控制 $P(\bar{X} \leq c) \leq \alpha$.

根据 CLT, 注意到

$$\bar{X} - \mu \sim N(0, \frac{\sigma^2}{n}) \Leftrightarrow \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1).$$

因此

$$\begin{aligned}P(\bar{X} \leq c) &= P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{c - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \\&= P\left(Z \leq \frac{c - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \\&= \Phi\left(\frac{c - \mu}{\frac{\sigma}{\sqrt{n}}}\right), \forall \mu \geq \mu_0.\end{aligned}$$

取 $\frac{c - \mu_0}{\frac{\sigma}{\sqrt{n}}} = -Z_\alpha \Rightarrow c = \mu_0 - \frac{\sigma}{\sqrt{n}} \cdot Z_\alpha$.

给出在 α 的检验水平下拒绝 H_0 的条件: 若 $\bar{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} \cdot Z_\alpha$, 则拒绝 H_0 .

- 注

- 这种方法称为 Z -检验.
- 若 σ^2 未知, 考虑

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

这种方法称为 t -检验.

2 Neyman-Pearson 假设检验

- 定义(错误) I 类错误: 在 H_0 为真时拒绝 H_0 ; II 类错误: 在 H_0 为假时不拒绝 H_0 . 对应的概率分别为:

$$\begin{aligned}\alpha(R) &:= P_\theta(\text{I}) = P_\theta((X_1, \dots, X_n) \in R), \theta \in \Theta_0; \\ \beta(R) &:= P_\theta(\text{II}) = P_\theta((X_1, \dots, X_n) \in R^c), \theta \in \Theta_1.\end{aligned}$$

对于已划分的 R 来说, 是定义域不同的 θ 的函数.

- 注

- 依据样本做决策, 错误不能根本避免;
- 一次决策不能同时犯两种错误;
- n 固定, 两种错误发生的概率此消彼长.

- 例(直觉)

- 都不拒绝 H_0 , 那么 $P_\theta(\text{I}) = 0, P_\theta(\text{II}) = 1$.
- 考虑事件 H_0 : 合格, 事件 H_1 : 不合格, 当 $P_\theta(\text{I})$ 变小时, 不容易拒绝事件, 不合格不容易被检出, 从而 $P_\theta(\text{II})$ 变大.

- 定义(功效函数) 给定 θ 与临界域 R , 拒绝原假设 H_0 的概率为:

$$P_\theta((X_1, \dots, X_n) \in R) = 1 - \beta(R), \theta \in \Theta_1.$$

- **定义(Neyman-Pearson 范式)** n 固定, 控制 $P_{\theta}(I) \leq \alpha$, 其中 α 为预先给定的检验水平, 再在这个限制下使 $P_{\theta}(II)(\theta \in \Theta_1)$ 尽可能小.
- **注**
 - α 固定时, 使 $P_{\theta}(II)(\theta \in \Theta_1)$ 最小的检验称为水平 α 下的一致最优检验;
 - 原假设 H_0 和备择假设 H_1 一般是地位不对等的:
 - 原假设通常是受到保护的, 证据不充分不能拒绝;
 - 备择假设通常是真正感兴趣的.
 - 一致最优检验不一定存在, 一般也不易求解;
 - $\mu_0 \in$ 置信区间 \Leftrightarrow 假设检验 ($H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$) 不拒绝 H_0 .

3 假设检验与置信区间

- **例** $X \sim N(\mu, \sigma^2)$, σ^2 已知, $\alpha > 0$ 给定, X_1, \dots, X_n 为随机样本.
- **解答** 其双侧置信区间为

$$P = (\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}).$$

考虑假设检验 $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$. 需要控制

$$P_{H_0}(|\bar{X} - \mu_0| \geq c) \leq \alpha.$$

检验准则为 $|\bar{X} - \mu_0| \geq Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ 时拒绝 H_0 . 我们给出接受域

$$\begin{aligned} R^c &= \{(X_1, \dots, X_n) \mid |\bar{X} - \mu_0| < Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\} \\ &= \{(X_1, \dots, X_n) \mid \mu_0 \in P\}. \end{aligned}$$

由此可见, $\mu_0 \in P \Leftrightarrow$ 用 \bar{X} 为检验统计量, 假设检验不拒绝 H_0 . (对偶关系)

4 检验的 P 值

- **定义** 当原假设 H_0 为真时, 检验统计量的观测值以及更极端观测出现的概率.
- **例(选举问题)** $n = 1200$, 调查到的支持比例为 $\frac{684}{1200} \approx 0.57$ (观测值).
- **解答** (1) $H_0: p = p_0$ v.s. $H_1: p > p_0$ (p_0 预先给定).

考虑检验统计量 P_n , 则由 CLT:

$$\frac{P_n - p}{\sqrt{\frac{p(1-p)}{n}}} \stackrel{\text{近似}}{\sim} N(0, 1).$$

当 H_0 为真时, $p = p_0$, 此时 $se(P_n) = \sqrt{\frac{p_0(1-p_0)}{n}}$. 因此 P 值

$$P_{H_0}(P_n \geq p_n) = P\left(\frac{P_n - p_0}{se(P_n)} \geq \frac{p_n - p_0}{se(P_n)}\right) \approx P(Z \geq z_0).$$

其中

$$Z \stackrel{\text{近似}}{\sim} N(0, 1), z_0 = \frac{p_n - p_0}{se(P_n)} = \frac{p_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

因此

$$P_{H_0}(P_n \geq p_n) = 1 - \Phi\left(\frac{p_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right).$$

若 $p_0 = 0.55$, 则 **P** 值 ≈ 0.081 ; 若 $p_0 = 0.5$, 则 **P** 值 $\ll 0.001$.

在水平 α 下拒绝 $H_0 \Leftrightarrow \mathbf{P}$ 值 $\leq \alpha$.

• 注

- **P** 值作为数据 (观测) 拒绝 H_0 的证据.
- 强弱的度量: **P** 值越小, 拒绝 H_0 的证据越强 (非正式的).
- **P** 值 $\neq P(H_0 \mid \text{观测})$.
- 若 **P** 值不小, 则不拒绝 H_0 , 原因可能为 H_0 真 / H_0 不真, 但检验功效不大.

- **定义** 若拒绝 $H_0 : \theta \in \Theta_0 \Leftrightarrow T(X_1, \dots, X_n) \geq C$, 则检验的 **P** 值 $:=$

$$\sup_{\theta \in \Theta_0} P_\theta(T(X_1, \dots, X_n) \geq T(x_1, \dots, x_n)).$$

其中 $T(x_1, \dots, x_n)$ 为检验统计量的观测值.

- **解答** (2) $H_0 : p \leq p_0$ v.s. $H_1 : p > p_0$ (p_0 预先给定).

考虑检验统计量 P_n , 则由 **CLT**:

$$\frac{P_n - p}{\sqrt{\frac{p(1-p)}{n}}} \stackrel{\text{近似}}{\sim} N(0, 1).$$

当 H_0 为真时, 此时 $se(P_n) \approx \hat{se}(P_n) = \sqrt{\frac{p_n(1-p_n)}{n}}$. 因此 **P** 值

$$P_{H_0}(P_n \geq p_n) = P\left(\frac{P_n - p}{\hat{se}(P_n)} \geq \frac{p_n - p}{\hat{se}(P_n)}\right) \approx P(Z \geq z_0).$$

其中

$$Z \stackrel{\text{近似}}{\sim} N(0, 1), z_0 = \frac{p_n - p}{\hat{se}(P_n)} = \frac{p_n - p}{\sqrt{\frac{p_n(1-p_n)}{n}}}.$$

因此

$$P_{H_0}(P_n \geq p_n) = 1 - \Phi\left(\frac{p_n - p}{\sqrt{\frac{p_n(1-p_n)}{n}}}\right), \forall p \leq p_0.$$

因此 **P** 值

$$\sup_{p \leq p_0} \left(1 - \Phi\left(\frac{p_n - p}{\sqrt{\frac{p_n(1-p_n)}{n}}}\right)\right) = 1 - \Phi\left(\frac{p_n - p_0}{\sqrt{\frac{p_n(1-p_n)}{n}}}\right).$$

若 $p_0 = 0.55$, 则 **P** 值 ≈ 0.081 ; 若 $p_0 = 0.5$, 则 **P** 值 $\ll 0.001$.

5 拟合优度检验

- 定义(Pearson 卡方统计量)

$$\chi^2 := \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \overset{\text{近似}}{\sim} \chi^2(k-1).$$

其中 O_i 为观测频数, E_i 为期望频数 (H_0 真的条件下).

- 定理 $H_0 : P(X \in \text{第 } i \text{ 单元}) = p_i (1 \leq i \leq k)$. 若 H_0 为真, 当 $n \rightarrow \infty$ 时, 有

$$\chi^2 \rightarrow \chi^2(k-1).$$

- 例 投掷一枚骰子 60 次.

点数	1	2	3	4	5	6	总计
观测频数	4	6	17	16	8	9	60
期望频数	10	10	10	10	10	10	60

H_0 : 分布均匀, H_1 : 分布不均匀.

- 解答 检验统计量的观测值:

$$\frac{(4 - 10)^2}{10} + \frac{(6 - 10)^2}{10} + \dots + \frac{(9 - 10)^2}{10} = 14.2.$$

自由度为 $6 - 1 = 5$. **P** 值 $= P_{H_0}(\chi^2 \geq 14.2) \approx 0.014$.

- 注 在实际应用中, 需要满足 $E_i = nP_i \geq 5$, 才能较好使用 **Pearson** 定理.
- 例(列联表独立性检验) 对某项议题态度与年龄段是否独立.

	青年	中年	老年	
支持	20	40	20	80
反对	30	30	10	70
	50	70	30	150

H_0 : 独立, H_1 : 不独立.

- 解答 $P_{ij} = P_{i+}P_{+j}$, 其中 P_{i+} , P_{+j} 称为边际概率.

在 H_0 为真前提下估计 P_{ij} . MLE:

$$P_{ij}^* = (P_{i+}P_{+j})^* = P_{i+}^* P_{+j}^* = \frac{\text{sum}(\text{row}_i)}{n} \times \frac{\text{sum}(\text{column}_j)}{n}.$$

得到

$$E_{ij} = nP_{ij} \approx nP_{ij}^* = \frac{1}{n} \text{sum}(\text{row}_i) \times \text{sum}(\text{column}_j).$$

计算得检验统计量观测值为 6.12, 自由度为 $(a-1)(b-1) = 2$.

得到 P 值 = $P_{H_0}(\chi^2 \geq 6.12) \approx 0.0469$.

- 注 对于 a 行 b 列, 当 H_0 成立时, 其未知参数个数为 $s = (a-1) + (b-1)$.

因此卡方自由度为:

$$ab - 1 - s = ab - 1 - (a-1) - (b-1) = (a-1)(b-1).$$

6 似然比检验

- 例 两种硬币 ($p = 0.5$, $p = 0.7$) 投掷 $n = 10$ 次, 正面向上 $X = x$ 次. 提出假设 $H_0: p = 0.5$ v.s. $p = 0.7$. 考虑

$$\frac{P_{H_0}(X=x)}{P_{H_1}(X=x)} \leq c \leftrightarrow x \text{ 的范围}$$

我们需要控制

$$P_{H_0}(\text{似然比} \leq c) \leq \alpha.$$

- 注
 - 当 H_0 , H_1 均为简单假设时 (N-P), 证明: 似然比检验最优 (功效最大).
 - 当 H_0 , H_1 不全为简单假设时, 似然比检验一般不最优, 但通常表现不错.
- 定义(广义似然比) $H_0: \theta \in \Theta_0$ v.s. $H_1: \theta \in \Theta_1$, X_1, \dots, X_n 为随机样本.

考虑广义似然比

$$\Lambda^* := \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta_1} L(\theta)}$$

基于技术原因, 检验统计量选为

$$\Lambda := \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta)} = \min\{\Lambda^*, 1\} = \Lambda(X_1, \dots, X_n).$$

Λ 越小则越反对 H_0 (拒绝域的形状). 选择 λ_0 使

$$P_{H_0}(\Lambda \leq \lambda_0) \leq \alpha.$$

至此得到了检验准则.

- **定理** 在一定(光滑性)条件下, 当 $n \rightarrow \infty$ 时, 在 H_0 为真前提下:

$$-2 \log \Lambda \rightarrow \chi^2(d).$$

其中自由度 $d = \dim(\Theta_0 \cup \Theta_1) - \dim(\Theta_0)$. 这里的 \dim 指自由参数的个数.

- **例(多项分布检验)** $H_0 : p_1 = p_1^*, \dots, p_k = p_k^*$, 观测频数分别为 n_1, \dots, n_k .
- **解答** 得到

$$L(p_1, \dots, p_k) = C_n^{n_1, \dots, n_k} p_1^{n_1} \dots p_k^{n_k}.$$

因此

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta)} = \frac{L(p_1^0, \dots, p_k^0)}{L(p_1^*, \dots, p_k^*)}.$$

已知

$$n_i = np_i^*, E_i = np_i^0, O_i = n_i.$$

且

$$x \log \frac{x}{x_0} \stackrel{Taylor}{=} 0 + (x - x_0) + \frac{1}{2} \frac{(x - x_0)^2}{x_0} + \dots$$

计算得

$$\begin{aligned}
-2 \log \Lambda &= -2 \sum_{i=1}^k \log \left(\frac{p_i^0}{p_i^*} \right)^{n_i} \\
&= -2 \sum_{i=1}^k n_i \log \frac{p_i^0}{p_i^*} \\
&= 2 \sum_{i=1}^k O_i \log \frac{O_i}{E_i} \\
&= \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} + \dots
\end{aligned}$$

考虑到 $\dim(\Theta_0) = 0$, $\dim(\Theta_0 \cup \Theta_1) = k - 1$, 因此给出

$$\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \underset{\text{近似}}{\sim} \chi^2(k - 1).$$

7 两独立总体比较

- 两独立总体:

总体	均值	方差	样本 (iid)
X	μ_1	σ_1^2	X_1, \dots, X_n
Y	μ_2	σ_2^2	Y_1, \dots, Y_m

- 定义(比较均值):

给出 $E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$, $Var(\bar{X} - \bar{Y}) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} = se^2$. 参考置信区间部分.

- 定理 给出 $W_1 \sim \chi^2(k_1)$, $W_2 \sim \chi^2(k_2)$, 且 W_1, W_2 独立, 我们有

$$\frac{W_1/k_1}{W_2/k_2} \sim F(k_1, k_2).$$

- 定义(比较方差): 假设 X, Y 均为正态. 提出假设 $H_0 : \sigma_1^2 = \sigma_2^2$ v.s. $H_1 : \sigma_1^2 \neq \sigma_2^2$.

$$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1), \frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1).$$

考虑检验统计量(依赖于样本或已知参数)

$$\frac{S_1^2}{S_2^2}.$$

当 H_0 为真时, 有

$$\frac{S_1^2}{S_2^2} \sim F(n-1, m-1).$$

检验准则为当

$$\frac{S_1^2}{S_2^2} \geq F_{\frac{\alpha}{2}}(n-1, m-1) \text{ 或 } \frac{S_1^2}{S_2^2} \leq F_{1-\frac{\alpha}{2}}(n-1, m-1)$$

时拒绝原假设.

- 注 由定义 $F_{1-\frac{\alpha}{2}}(n-1, m-1) = \frac{1}{F_{\frac{\alpha}{2}}(m-1, n-1)}$.
- 例(比较成功率/失败率) 阿司匹林对降低心脏病发病率的有效性.

	心脏病发作	心脏病未发作	合计	发作率
阿司匹林	139	10898	11037	0.0126
安慰剂	239	10795	11034	0.0217

提出假设 $H_0: p_1 = p_2$ (无效) v.s. $H_1: p_1 < p_2$ (有效), 检验统计量为 $P_1 - P_2$.

容易得到 $E(P_1 - P_2) = p_1 - p_2$, $Var(P_1 - P_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$.

那么

$$\frac{(P_1 - P_2) - (p_1 - p_2)}{se} \overset{\text{近似}}{\sim} N(0, 1).$$

在 H_0 为真前提下, 有

$$se^2 = p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \approx p^*(1-p^*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

其中 $p^* = \frac{k_1 + k_2}{n_1 + n_2}$, 得 $se^2 \approx \hat{se}^2 = 0.00175^2$. 结合 $\frac{P_1 - P_2}{\hat{se}} \overset{\text{近似}}{\sim} N(0, 1)$, 得 P 值

$$P\left(\frac{P_1 - P_2}{\hat{se}} \leq \frac{0.0126 - 0.0217}{0.00175}\right) \approx P(Z \leq -5.20) \approx 10^{-7}.$$

因此有理由拒绝 H_0 .

- 注
 - 随机分组.
 - 双盲实验.
 - n 充分大.
- 例(行驶里程) 比较两种油 A 与 B 的行驶里程.

	样本容量	平均里程	样本标准差
油 A	50	25	5.00
油 B	50	26	4.00

提出假设 $H_0 : \mu_A = \mu_B$ v.s. $H_1 : \mu_A \neq \mu_B$, 检验统计量为 $\bar{X}_A - \bar{X}_B$.
在 H_0 为真前提下, 有

$$\frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0, 1).$$

其中 $\hat{se} \approx 0.905$. 得 **P** 值

$$P\left(\left|\frac{\bar{X}_A - \bar{X}_B}{\hat{se}}\right| \geq \left|\frac{25 - 26}{0.905}\right|\right) \approx P(|Z| \geq 1.1) \approx 0.2714.$$

认为哪种油行驶里程更长的理由均不充分.

8 两相关总体比较

- 例(行驶里程-改进) 同一辆车不同日子加不同油, 记录行驶里程.

车号	油 A	油 B	差异 (d_i)
1	27.01	26.95	0.06
2	20.00	20.44	-0.44
...
10	25.22	26.01	-0.79
均值	25.20	25.80	-0.60
标准差	4.27	4.10	0.61

提出假设 $H_0 : \mu_d = 0$ v.s. $H_1 : \mu_d \neq 0$, 检验统计量为 \bar{d} .
假设 d_i 服从正态分布. 在 H_0 为真前提下, 有

$$\frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} \sim t(n - 1).$$

得 **P** 值

$$P\left(|t(9)| \geq \left|\frac{-0.60}{\frac{0.61}{\sqrt{10}}}\right|\right) \approx 0.012.$$

有理由拒绝 H_0 , 两种油的行驶里程有差距.

- 注 假设检验不能检验试验设计, 仅对数据负责; 功能有限, 作为决策辅助.

9 Bayes 假设检验

- 例 两种硬币 ($p = 0.5$, $p = 0.7$) 投掷 $n = 10$ 次, 正面向上 $X = x$ 次. 提出假设 $H_0 : p = 0.5$ v.s. $p = 0.7$. 考虑

$$\frac{P(H_0 | x)}{P(H_1 | x)} = \frac{P(H_0)P(x | H_0)}{P(H_1)P(x | H_1)} < 1$$

则拒绝 H_0 .

- 注 给出一个特别的例子, $H_0 : \theta = \theta_0$, Θ 连续. 则 $P(H_0 | x) = 0$, 此时似乎总是会拒绝 H_0 ? (陈书 Chap 5.28)

10 Review

10.1 决策

- 拒绝 H_0 或不拒绝 H_0 .
- 检验=决策准则 \Leftrightarrow 拒绝域 R 的划分.
- 关键:
 - 选择合适的检验统计量.
 - 确定拒绝域的形状 (由 H_1 决定).
- 拒绝 H_0 有时也称观测值是显著的.

10.2 错误

- 统计学中没有绝对的证实或证伪.

$$\begin{aligned}\alpha(R) &:= P_\theta((X_1, \dots, X_n) \in R | H_0). \\ \beta(R) &:= P_\theta((X_1, \dots, X_n) \in R^c | H_1).\end{aligned}$$

- 检验程序的属性, 不是样本的属性. 样本做决策要么正确要么错误.

$$\begin{aligned}\alpha(R) &\leq \alpha. \\ \beta(R) &\leq \beta.\end{aligned}$$

预先指定的可接受的长期错误率.

10.3 显著性检验 v.s. Neyman-Pearson 检验

- 显著性检验: 只控制 $\alpha(R) \leq \alpha$.
- Neyman-Pearson 假设检验: 强调两类错误、功效, H_0, H_1 地位不均等.
- 不拒绝 $H_0 \neq$ 接受 H_0 .
- $\beta(R)$ 越小 (功效越大), 当 H_0 不真时, 越有可能拒绝 H_0 ; 当观测支持 H_0 , 则可以接受 H_0 .
- 若忽略了对 $\beta(R)$ 的系统控制 (常见情形), 将导致对结果及下一步工作方向的误判.
- 例 $H_0: \mu \geq 5, H_1: \mu < 5. n = 10, \sigma = 0.01, \alpha = 0.01$.
- 解答 临界值

$$c = \mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}} \approx 4.993.$$

10.4 P 值

- 一次具体的观测值没有概率可言, **P** 不能衡量决策错误的概率. (**ASA** 文章)

10.5 卡方检验 —— 多项分布的检验

- 例 $H_0: p_1 = p, p_2 = 1 - p$. 此时

$$\begin{aligned}\chi^2 &= \frac{(O_1 - np_1)^2}{np_1} + \frac{(O_2 - np_2)^2}{np_2} \\ &= \frac{(O_1 - np)^2}{np} + \frac{(O_1 - np)^2}{n(1-p)} \\ &= \frac{(O_1 - np)^2}{np(1-p)} \\ &\stackrel{\text{近似}}{\sim} N^2(0, 1) = \chi^2(1).\end{aligned}$$

10.6 统计显著 \neq 实际显著

- 例 投掷骰子 $n = 6 \times 10^{10}$ 次.

点数	1	2	3	4	5	6	总计
观测频数 10^{10}	-10^6	1.5×10^6	-2×10^6	4×10^6	-3×10^6	0.5×10^6	6×10^{10}

H_0 : 分布均匀, H_1 : 分布不均匀.

- 计算得到 $\chi^2 = 3250$, 此时 **P** 值 $\ll 0.0001$. 因此拒绝 H_0 , 统计显著.
- 实际上 $|\hat{p}_i - \frac{1}{6}| \sim 10^{-4}$, 实际角度视为无差异, 实际不显著.
- n 过大, 明察秋毫; χ^2 统计量关于 n 是非齐次的.