

10 假设检验(Hypothesis Testing)

根据样本信息来检验关于总体的某个假设是否正确, 此类问题称为 **假设检验问题**, 可分为两类:

- 参数检验问题: 总体分布已知, 检验某未知参数的假设;
- 非参数检验问题: 总体分布未知时的假设检验问题.

假设检验的方法: 先假设所做的假设 H_0 成立, 然后从总体中取样, 根据样本的取值来判断是否有‘不合理’的现象出现, 最后做出接受或者拒绝所做假设的决定. ‘不合理’的现象指小概率事件在一次事件中几乎不会发生.

例10.1. 某产品出厂检验规定次品率 $p \leq 0.04$ 才能出厂, 现从 10000 件产品中任抽取 12 件, 发现 3 件是次品, 问该批产品是否该出厂; 若抽样结果有 1 件次品, 问该批产品是否该出厂?

解. 首先做出假设 $H_0: p \leq 0.04$. 若假设 H_0 成立, 设随机变量 $X \sim B(12, p)$,

$$\Pr[X = 3] = \binom{12}{3} p^3 (1-p)^9 \leq 0.0097.$$

由此可知这是一个小概率事件, 一次试验不应该发生, 但却发生了, 故不合理, 原假设 $H_0: p \leq 0.04$ 不成立, 即 $p > 0.04$, 该批产品不能出厂.

若 $X = 1$ 则

$$\Pr[X = 1] = p(1-p)^{11} \binom{12}{1} \geq 0.306.$$

这不是小概率事件, 没理由拒绝原假设 H_0 , 产品可以出厂. □

注: 当 $X = 1$ 情况下, 若直接利用参数估计

$$p = 1/12 = 0.083 > 0.04.$$

若仅仅采用参数估计而不用假设检验, 则不能出厂, 因此参数估计与假设检验是两回事.

在假设检验中, 需要对‘不合理’的小事件给出一个定性描述, 通常给出一上界 α , 当一事件发生的概率小于 α 时则成为小概率事件. 通常取 $\alpha = 0.05, 0.1, 0.01$, 其具体取值根据实际问题而定. 在假定 H_0 成立下, 根据样本提供的信息判断出不合理现象 (概率小于 α 的事件发生), 则认为假设 H_0 不显著, α 被称为显著水平.

注意: 不否定假设 H_0 并不是肯定假设 H_0 一定成立, 而只能说差异不够显著, 没达到否定的程度, 所以假设检验被称为“显著性检验”.

前面的例子初步介绍了假设检验的基本思想和方法, 下面再进一步说明假设检验的一般步骤:

例10.2. 假设某产品的重量服从 $\mathcal{N}(500, 16)$, 随机取出 5 件产品, 测得重量为 509, 507, 498, 502, 508, 问产品的期望是否正常? (显著性水平 $\alpha = 0.05$)

解. 下面给出假设检验的一般步骤:

- 第一步: 提出原假设 $H_0: \mu = 500$ 和备择假设 $H_1: \mu \neq 500$;
- 第二步: 设计检验统计量, 在原假设 H_0 成立下的条件下求出其分布. 令样本均值 $\bar{X} = \sum_{i=1}^5 X_i/5 = 504.8$, 设检验统计量为

$$Z = \frac{\bar{X} - 500}{\sqrt{16/5}} \sim \mathcal{N}(0, 1).$$

检验统计量能衡量差异大小且分布已知.

- 第三步: 给定显著性水平 $\alpha = 0.05$, 查表得到临界值 $\mu_{0.025} = 1.96$, 使得

$$\Pr[|Z| > 1.96] = 0.05$$

成为一个小事件, 从而得到否定域 $\{Z: |Z| > 1.96\}$.

- 第四步: 将样本值代入计算统计量 Z 的实测值

$$|Z| = \frac{|\bar{X} - 500|}{\sqrt{16/5}} = \frac{4.8}{4/\sqrt{5}} = 1.2 \times \sqrt{5} = 2.68 > 1.96.$$

根据实测值 Z 落入否定域 $\{Z: |Z| > 1.96\}$, 从而拒绝原假设 H_0 .

□

由此归纳出假设检验的一般步骤:

- 1) 根据实际问题提出原假设 H_0 和备择假设 H_1 ;
- 2) 确定检验统计量 (分布已知);
- 3) 确定显著性水平 α , 并给出拒绝域;
- 4) 由样本计算统计量的实测值, 判断是否接受原假设 H_0 .

假设检验可分为如下三类:

- 原假设 $H_0: \mu = \mu_0$ 和备选假设 $H_1: \mu \neq \mu_0$, 称为 **双边假设检验**;
- 原假设 $H_0: \mu \leq \mu_0$ 和备选假设 $H_1: \mu > \mu_0$, 称为 **右边检验**;
- 原假设 $H_0: \mu \geq \mu_0$ 和备选假设 $H_1: \mu < \mu_0$, 称为 **左边检验**.

右边检验和左边检验又被通称为双边检验.

下面研究假设检验是否会犯错, 假设检验的核心是先假设原判断假设 H_0 成立, 然后根据样本的取值来判断是否有‘不合理’的现象出现, 即“小概率”原理, 然而小概率事件在一次试验中不发生并不意味着小概率事件不发生. 可能发生如下两种错误:

- 第 I 类错误: “弃真”, 即当 H_0 为真时, 我们仍可能拒绝 H_0 .
- 第 II 类错误: “存伪”, 即当 H_0 不成立时, 我们仍可能接受 H_0 .

两类错误如下表格所示

假设检验的决定	真实情况: H_0 为真	真实情况: H_0 为假
拒绝 H_0	第 I 类错误	正确
接受 H_0	正确	第 II 类错误

设犯第 I 类错误的概率为 α , 即显著性水平, 第 II 类错误的概率用 β 表示, 即

$$\alpha = \Pr[\text{拒绝 } H_0 | H_0 \text{ 为真}] \quad \beta = \Pr[\text{接受 } H_0 | H_0 \text{ 为假}].$$

这两类错误互相关联, 当样本容量固定时, 一类错误概率的减少导致另一类错误概率的增加. Neyman-Pearson 原则: 在控制第 I 类错误的前提下, 尽可能减小第 II 类错误的概率.

10.1 正态总体期望的假设检验

10.1.1 方差已知的单个正态总体的期望检验 (Z 检验)

设 X_1, X_2, \dots, X_n 是来自总体 $X \sim N(\mu, \sigma^2)$ 的样本, 若方差 σ^2 已知, 检验原假设 $H_0: \mu = \mu_0$ 和备择假设 $H_1: \mu \neq \mu_0$. 设样本均值为 $\bar{X} = \sum_{i=1}^n X_i/n$, 根据正态分布的性质选择检验统计量

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

给定显著性水平 α , 得到拒绝域为 $|Z| \geq \mu_{\alpha/2}$, 这种检验方法称为 **Z 检验法**.

关于 Z 检验法的双边和单边检验有

- 原假设 $H_0: \mu = \mu_0$ 和备择假设 $H_1: \mu \neq \mu_0$, 拒绝域为 $\{Z: |Z| \geq \mu_{\alpha/2}\}$;
- 原假设 $H_0: \mu \geq \mu_0$ 和备择假设 $H_1: \mu < \mu_0$, 拒绝域为 $\{Z: Z \leq -\mu_{\alpha}\}$;
- 原假设 $H_0: \mu \leq \mu_0$ 和备择假设 $H_1: \mu > \mu_0$, 拒绝域为 $\{Z: Z \geq \mu_{\alpha}\}$.

例10.3. 已知某产品的重量 $X \sim \mathcal{N}(4.55, 0.108^2)$, 现随机抽取 5 个产品, 其质量分别为 4.28, 4.40, 4.42, 4.35, 4.27. 问产品的期望在 $\alpha = 0.05$ 下有无显著性变化. ($\mu_{0.025} = 1.96$)

解. 首先提出原假设 $H_0: \mu = 4.55$ 和备择假设 $H_1: \mu \neq 4.55$. 若 H_0 成立, 选择检验量

$$Z = \frac{\bar{X} - 4.55}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

求得拒绝域为 $|Z| \geq \mu_{\alpha/2} = 1.96$. 计算样本均值可知 $\bar{X} = 4.364$, 于是有

$$\frac{\bar{X} - 4.55}{0.108/\sqrt{5}} = 3.851 > 1.96,$$

由此可拒绝 H_0 , 说明有显著变化. □

例10.4. 某灯泡平均寿命要求不低于 1000 小时被称为‘合格’, 已知灯泡的寿命 $X \sim \mathcal{N}(\mu, 100^2)$, 现在随机抽取 25 件, 其样本均值为 $\bar{X} = 960$. 在显著性水平 $\alpha = 0.05$ 的情况下, 检验这批灯泡是否合格. ($\mu_{0.05} = 1.645$)

解. 首先提出原假设 $H_0: \mu \geq 1000$ 和备择假设 $H_1: \mu < 1000$. 若 H_0 成立, 选择假设统计量

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

由此得到假设拒绝域为: $Z < -\mu_\alpha = -1.645$. 根据样本均值 $\bar{X} = 960$ 可知观察值

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = -2.0 < -1.645$$

由此可拒绝 H_0 , 认为这篇灯泡不合格. □

10.1.2 方差未知的单个正态总体的期望检验 (t 检验)

设 X_1, X_2, \dots, X_n 是来自总体 $X \sim N(\mu, \sigma^2)$ 的样本, 若方差 σ^2 未知, 检验原假设 $H_0: \mu = \mu_0$ 和备择假设 $H_1: \mu \neq \mu_0$. 设样本均值为 $\bar{X} = \sum_{i=1}^n X_i/n$ 和样本修正方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, 根据正态分布的性质选择检验统计量

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

给定显著性水平 α , 得到拒绝域为 $|t| \geq t_{\alpha/2}(n-1)$, 这种检验方法称为 **t 检验法**.

关于 t 检验法的双边和单边检验有

- i) 原假设 $H_0: \mu = \mu_0$ 和备择假设 $H_1: \mu \neq \mu_0$, 拒绝域为 $\{t: |t| \geq t_{\alpha/2}(n-1)\}$;
- ii) 原假设 $H_0: \mu \geq \mu_0$ 和备择假设 $H_1: \mu < \mu_0$, 拒绝域为 $\{t: t \leq -t_\alpha(n-1)\}$;
- iii) 原假设 $H_0: \mu \leq \mu_0$ 和备择假设 $H_1: \mu > \mu_0$, 拒绝域为 $\{t: t \geq t_\alpha(n-1)\}$.

10.1.3 方差已知的两个正态总体的期望差检验

设 X_1, X_2, \dots, X_n 是来自总体 $X \sim N(\mu_1, \sigma_1^2)$ 的样本, 以及 Y_1, Y_2, \dots, Y_m 是来自总体 $Y \sim N(\mu_2, \sigma_2^2)$ 的样本, 若方差 σ_1^2 和 σ_2^2 已知, 检验原假设 $H_0: \mu_1 - \mu_2 = \delta$ 和备择假设 $H_1: \mu_1 - \mu_2 \neq \delta$ (注: δ 为常数). 设样本均值 $\bar{X} = \sum_{i=1}^n X_i/n$ 和 $\bar{Y} = \sum_{i=1}^m Y_i/m$, 根据正态分布的性质有

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

给定显著性水平 α , 其双边和单边检验有

- i) 原假设 $H_0: \mu_1 - \mu_2 = \delta$ 和备择假设 $H_1: \mu_1 - \mu_2 \neq \delta$, 拒绝域为 $\{U: |U| \geq \mu_{\alpha/2}\}$;
- ii) 原假设 $H_0: \mu_1 - \mu_2 \geq \delta$ 和备择假设 $H_1: \mu_1 - \mu_2 < \delta$, 拒绝域为 $\{U: U \leq -\mu_\alpha\}$;
- iii) 原假设 $H_0: \mu_1 - \mu_2 \leq \delta$ 和备择假设 $H_1: \mu_1 - \mu_2 > \delta$, 拒绝域为 $\{U: U \geq \mu_\alpha\}$.

10.1.4 方差未知但相等的两个正态总体的期望差检验

略, 以后补上

10.1.5 基于成对 (pairwise) 数据的检验

在很多实际应用中, 为了比较两种方法或两种产品的差异, 往往会得到一批成对的观察值, 然后基于观察的数据分析判断方法或产品是否具有显著的区别, 这种方法称为 **成对 (pairwise) 比较法**.

例10.5. 假设有两种学习方法 A 和 B , 在 9 个数据集上取得的效果如下表

数据集	1	2	3	4	5	6	7	8	9
方法 A	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.9	0.7
方法 B	0.7	0.95	0.7	0.6	0.7	0.9	0.9	0.8	0.6

问这两种方法在 $\alpha = 0.05$ 下是否有显著性区别?

上述问题可进一步形式化为: 假设观察到 n 对互相独立的随机变量 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, 其中 X_1, X_2, \dots, X_n 和 Y_1, Y_2, \dots, Y_n 分别是总体 X 和 Y 的两个样本, 检验这两种方法是否性能相同, 即检验总体 X 和 Y 的期望是否相等. 因为对相同的数据集 i 而言, X_i 和 Y_i 不能被认为相互独立. 由此假设

$$Z = X - Y \sim \mathcal{N}(\mu, \sigma^2),$$

并提出原假设 $H_0: \mu = 0$ 和备择假设 $H_1: \mu \neq 0$, 方差 σ^2 未知, 因此考虑统计 t 检验量. 设 $Z_i = X_i - Y_i$ ($i \in [n]$), 可得样本均值和方差分别为

$$\bar{Z} = \sum_{i=1}^n \frac{Z_i}{n} \quad \text{和} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2.$$

由此得到统计检验量

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

在显著性水平 α 下得到拒绝域为: $|t| > t_{\alpha/2}(n-1)$. 下面给出例 10.5 详细求解.

解. 设随机变量 $Z_i = X_i - Y_i$ ($i \in [10]$), 可得样本均值 $\bar{Z} = 0.0056$ 和方差 $S^2 = 0.009$, 由此可得观察值

$$|t| = \frac{|\bar{Z}|}{S/\sqrt{n}} = \frac{0.0056}{0.9} \approx 0.062 < t_{0.025}(8) = 2.3060,$$

由此说明这两种方法没有显著性区别. □

10.2 正态分布的方差假设检验.

10.2.1 单个正态总体的方差检验 (χ^2 检验)

设 X_1, X_2, \dots, X_n 是来自总体 $X \sim N(\mu, \sigma^2)$ 的样本, 检验原假设 $H_0: \sigma^2 = \sigma_0^2$ 和备择假设 $H_1: \sigma^2 \neq \sigma_0^2$. 设样本修正方差 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$, 根据正态总体抽样定理选择检验统

计量

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

给定显著性水平 α 求解拒绝域, 这种检验方法称为 χ^2 检验法.

关于 χ^2 检验法的双边和单边检验有

- i) 原假设 $H_0: \sigma^2 = \sigma_0^2$ 和备择假设 $H_1: \sigma^2 \neq \sigma_0^2$, 拒绝域为: $\{\chi^2 \geq \chi_{\frac{\alpha}{2}}^2(n-1)\} \cup \{\chi^2 \leq \chi_{1-\frac{\alpha}{2}}^2(n-1)\}$.
- ii) 原假设 $H_0: \sigma^2 \geq \sigma_0^2$ 和备择假设 $H_1: \sigma^2 < \sigma_0^2$. 拒绝域为: $\{\chi^2 \leq \chi_{1-\alpha}^2(n-1)\}$.
- iii) 原假设 $H_0: \sigma^2 \leq \sigma_0^2$ 和备择假设 $H_1: \sigma^2 > \sigma_0^2$. 拒绝域为: $\{\chi^2 \geq \chi_{\alpha}^2(n-1)\}$.

10.2.2 两个正态总体的方差比检验 (F 检验)

略

10.3 非参假设检验

前面的内容讨论整体分布类型已知 (正态总体) 的参数假设检验问题. 本节讨论总体分布的假设检验问题, 因为所研究的检验是如何利用子样去拟合总体分布, 所以又被称分布的拟合优度检验.

10.3.1 χ^2 检验法

设总体 X 的分布函数 $F(x)$ 具体形式未知. 根据样本 X_1, \dots, X_n 来检验关于总体的假设:

$$H_0: F(x) = F_0(x)$$

其中 $F_0(x)$ 为某确定的分布函数.

若总体 X 为离散随机变量: $H_0: \Pr[X = x_i] = p_i \ (i = 1, 2, \dots)$

若总体 X 为连续随机变量: $H_0: X$ 的密度函数 $p(x) = p_0(x)$

若 p_i 或 $p_0(x)$ 包含未知参数, 此时首先用极大似然估计/矩估计估计未知参数.

下面介绍 χ^2 检验法: 将随机试验结果的全体 Ω 分成 k 个互不相容的事件 A_1, A_2, \dots, A_k , 且 $\cup_{i=1}^k A_i = \Omega$. 根据假设 $H_0: F(x) = F_0(x)$ 计算概率 $p_i = \Pr(A_i)$. 对样本 X_1, \dots, X_n , 事件 A_i 出现的频率为 n_i/n . 当假设 H_0 为真时, 频率 n_i/n 与概率 p_i 差异不应太大. 基于这种思想, Pearson 构造了检验统计量:

$$W = \sum_{i=1}^K \frac{(n_i - np_i)^2}{np_i}$$

称为 Pearson χ^2 统计量.

定理10.1. 若分布函数 $F_0(x)$ 不包含未知参数, 当 H_0 为真时 (无论 H_0 中的分布属于什么分布), 统计量

$$W = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(k-1)$$

证明超出了本书的范围. 给定显著性水平 α , 若 $W > \chi_\alpha^2(k-1)$ 则拒绝 H_0 .

例10.6. 实验 E 有四种不同的结果 $\{A, B, C, D\}$. 现进行如下实验: 独立重复实验直到结果 A 发生为止. 记录下抛掷的次数, 如此试验 200 次, 结果如下表. 试问该试验是否为均匀分布?

重复次数	1	2	3	4	≥ 5
频数	56	48	32	28	36

解. 首先提出原假设 H_0 : 均匀分布. 用随机变量 X 表示试验结果 A 发生时重复的试验次数, 有

$$p_1 = P(X=1) = \frac{1}{4} \quad p_2 = P(X=2) = \frac{3}{4} \times \frac{1}{4} \quad p_3 = P(X=3) = \left(\frac{3}{4}\right)^2 \cdot \frac{1}{4}$$

$$p_4 = P(X=4) = \left(\frac{3}{4}\right)^3 \cdot \frac{1}{4} \quad p_5 = P(X=5) = 1 - \frac{1}{4} - \frac{3}{16} - \left(\frac{3}{4}\right)^3 \cdot \frac{1}{4}$$

计算检验统计量

$$W = \sum_{i=1}^5 \frac{(n_i - np_i)^2}{np_i} = 18.21$$

根据统计量实值 $W > \chi_{0.05}^2(4) = 9.488$, 因此不服从均匀分布. \square

上例指定了分布的具体分布形式. 在许多实际问题中, 假设 H_0 只确定了总体分布的类型, 分布中还包含未知参数, 如

$$H_0: F(x) = F_0(x; \theta_1, \theta_2, \dots, \theta_r)$$

其中 F_0 已知, $\theta_1, \theta_2, \dots, \theta_r$ 未知. 从样本 X_1, X_2, \dots, X_n 中得到估计值 $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r$, 代入得

$$H_0: F(x) = F_0(x; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r)$$

将子样分成 k 组: $a_0 < a_1 < \dots < a_k$ 且 $A_1 \in [a_0, a_1], A_2 \in [a_1, a_2] \dots A_k = [a_{k-1}, a_k]$. 总体 X 落入 A_i 的概率为

$$\hat{p}_i = p(x \in A_i | \hat{\theta}_1, \dots, \hat{\theta}_r)$$

检验估计量 W 为

$$W = \sum_{i=1}^k \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

定理10.2. 当 $n \rightarrow +\infty$ 时, 有 $W \xrightarrow{d} \chi^2(k-r-1)$ 成立.

10.3.2 独立性检验

设 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ 是总体 (X, Y) 的样本, 通过样本考虑二元总体 (X, Y) 中随机变量 X 与 Y 的独立性. 将随机变量 X 和 Y 的取值分成 r 个和 s 个互不相交的区间 A_1, A_2, \dots, A_r 和

B_1, B_2, \dots, B_s . 用 n_{ij} 表示落入区域 $A_i \times B_j$ 的频数. 设 $n_{i\cdot} = \sum_{j=1}^s n_{ij}$ 和 $n_{\cdot j} = \sum_{i=1}^r n_{ij}$ 为边缘之和, 则 $n = \sum_{i,j} n_{ij}$. 建立如下二元联立表:

	B_1	B_2	\cdots	B_s	$n_{i\cdot}$
A_1	n_{11}	n_{12}	\cdots	n_{1s}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	\cdots	n_{2s}	$n_{2\cdot}$
\vdots	\vdots	\vdots		\vdots	\vdots
A_r	n_{r1}	n_{r2}	\cdots	n_{rs}	$n_{r\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	\cdots	$n_{\cdot s}$	n

首先提出假设 H_0 : X 与 Y 相互独立. 记

$$p_{ij} = \Pr(X \in A_i, Y \in B_j) \quad p_{i\cdot} = P(X \in A_i) = \sum_{j=1}^s p_{ij} \quad p_{\cdot j} = P(Y \in B_j) = \sum_{i=1}^r p_{ij}$$

若假设 H_0 成立, 则 $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$. 利用矩估计/最大似然估计得

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}, \quad \hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}.$$

设计假设检验统计量

$$W = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}} = n \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - n \sim \chi^2((r-1)(s-1))$$

在显著性水平为 α 时有 $W \sim \chi^2((r-1)(s-1))$ 成立, 由此得到拒绝域为: $W > \chi_\alpha^2((r-1)(s-1))$, 即在此范围内不接受随机变量 X 与 Y 独立.