

Ch11: 假设检验

Hypothesis Testing

December 13, 2023

假设检验

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.

假设检验方法 (反证法):

- 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”的现象是指小概率事件在一次事件中几乎不会发生.

假设检验：例 0.152

例 0.152 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?

解答：例 0.152

题目：如上所述.

解答：

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

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

由此可知这是一个小概率事件, 即“不合理”的现象发生了, 应该拒绝原假设 $H_0 : p \leq 0.04$, 即 $p > 0.04$, 该产品不该出厂.

- 若 $X = 1$ 则

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

这不是一个小概率事件, 故没有理由拒绝原假设 H_0 , 产品可以出厂.

- 当 $X = 1$ 情况下, 若直接利用参数估计计算 $p = 1/12 = 0.083 > 0.04$, 则不能出厂, 因此参数估计与假设检验是两回事.

建立假设

定义 0.91 设来自某个参数分布 $F\{(x, \theta) | \theta \in \Theta\}$ 的样本 X_1, X_2, \dots, X_n , 其中 Θ 是参数空间, 设 $\Theta_0 \subset \Theta$, 且 $\Theta_0 \neq \emptyset$, 则命题 $H_0 : \theta \in \Theta_0$ 称为**原假设或零假设 (null hypothesis)**. 若有另一个 $\Theta_1 (\Theta_1 \subset \Theta, \Theta_0 \cap \Theta_1 = \emptyset)$, 则命题 $H_1 : \theta \in \Theta_1$ 称为**对立假设或备择假设 (alternative hypothesis)**. 记为

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1$$

Remarks:

- 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 $H_1 : \mu > \mu_0$, 称为单边 (右边) 检验
- 原假设 $H_0 : \mu \geq \mu_0$ 和备选假设 $H_1 : \mu < \mu_0$, 称为单边 (左边) 检验

选择检验统计量

由样本对原假设进行检验总是通过一个统计量完成的, 该统计量称为检验统计量.

- 总体均值的检验统计量可选为样本均值 \bar{X}
- 总体方差的检验统计量可选为无偏样本方差 S^2
- 事件 A 发生的概率的检验统计量可选为事件 A 出现的频率
- ...

确定显著水平

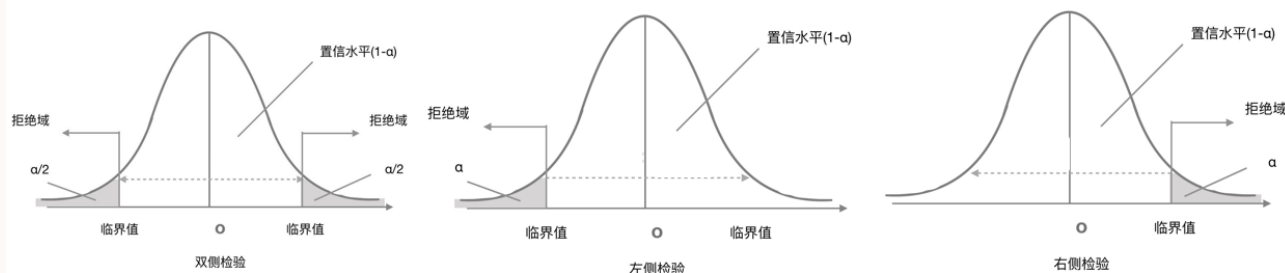
在假设检验中, 需要对“不合理”的事件给出一个定性描述: 即给出一个上界 α , 当一事件发生的概率小于 α 时, 称为小概率事件. 通常取 $\alpha = 0.05, 0.1, 0.01$, 其具体取值根据实际问题而定.

- 在假定 H_0 成立下, 根据样本提供的信息判断出不合理的现象 (即, 概率小于 α 的事件发生了), 则认为假设 H_0 不显著, α 被称为显著水平.
- 但是不否定假设 H_0 并不代表假设 H_0 一定成立, 而只能说试验结果与假设 H_0 之间的差异不够显著, 没达到否定的程度, 所以假设检验也被称为“显著性检验”.

给出拒绝域, 由样本统计量做判断

给定显著性水平 α 后, 查表就可以得到具体临界值, 拒绝域是由显著性水平围成的区域, 拒绝域通常记为 W .

- 拒绝域的功能主要用来判断假设检验是否拒绝原假设的.
- 由样本观测值 X_1, X_2, \dots, X_n 计算出来的检验统计量 $T(X_1, X_2, \dots, X_n)$ 判断是否拒绝原假设:
 - 如果 $T(X_1, X_2, \dots, X_n) \in W$, 则拒绝 H_0 .
 - 如果 $T(X_1, X_2, \dots, X_n) \in \overline{W}$, 则接受 H_0 .



检验的两类错误

我们通过样本数据来判断总体参数的假设是否成立, 但样本是随机的, 因而有可能出现小概率的错误. 这种错误分两种, 一种是弃真错误, 另一种是取伪错误.

- 第 I 类错误: “弃真”, 即当 H_0 为真时仍可能拒绝 H_0 . 设犯第 I 类错误的概率为 α , 即显著性水平, 则有:

$$\alpha = \Pr[\text{拒绝}H_0 \mid H_0\text{为真}]$$

- 第 II 类错误: “存伪”, 即当 H_0 不成立时仍可能接受 H_0 . 设犯第 II 类错误的概率为 β , 即显著性水平, 则有:

$$\beta = \Pr[\text{接受}H_0 \mid H_0\text{为假}]$$

检验的两类错误 – – Remarks

- 第 I 类错误与第 II 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. 既然我们不可能同时控制一个检验犯第 I 类错误与第 II 类错误的概率, 则采取折中方案 (即 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.
- 由于 Neyman-Pearson 原则提出了要控制犯第 I 类错误的概率 α , 因此在假设检验中, 通常将不宜轻易拒绝的假设作为原假设.

假设检验：例 0.153

例 0.153 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 $\alpha = 0.05$)

解答：例 0.153

解答：

- 建立假设： $H_0 : \mu = 500$ vs $H_1 : \mu \neq 500$.
- 设计检验统计量：在原假设 H_0 成立下的条件求出其分布. 令样本均值 $\bar{X} = \sum_{i=1}^5 X_i / 5 = 504.8$, 设检验统计量为

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

- 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, 从而得到拒绝域 $\{Z : |Z| > 1.96\}$.

- 将样本值代入计算统计量 Z 的实测值

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

由于实测值落入拒绝域, 因此判断为拒绝原假设 H_0 .

假设检验：例 0.154

例 0.154 设 (X_1, X_2, X_3, X_4) 是取自正态分布 $\mathcal{N}(\mu, 1)$ 的一个样本, 检验假设

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu = 1 ,$$

拒绝域为 $W = \{\bar{X} \geq 0.98\}$, 求此检验的两类错误概率.

解答：例 0.154

题目：设 (X_1, X_2, X_3, X_4) 是取自正态分布 $\mathcal{N}(\mu, 1)$ 的一个样本，检验假设 $H_0 : \mu = 0$ vs $H_1 : \mu = 1$ ，拒绝域为 $W = \{\bar{X} \geq 0.98\}$ ，求此检验的两类错误概率。

解答：

- 第 I 类错误是指原假设 H_0 成立时，但由于样本落入拒绝域而做出了拒绝原假设的情况，当原假设 H_0 成立时， $\bar{X} \sim \mathcal{N}(0, 1/4)$ ，因此犯第 I 类错误的概率为

$$P_{H_0}[\bar{X} \geq 0.98] = 1 - \Phi\left(\frac{0.98 - 0}{\sqrt{1/4}}\right) = 1 - \Phi(1.96) = 0.025.$$

- 第 II 类错误是指原假设 H_0 不成立而接受备择假设 H_1 时，但由于样本落入接受域而做出了不拒绝原假设的情况，当原假设 H_1 成立时， $\bar{X} \sim \mathcal{N}(1, 1/4)$ ，因此犯第 II 类错误的概率为

$$P_{H_1}[\bar{X} < 0.98] = 1 - \Phi\left(\frac{0.98 - 1}{\sqrt{1/4}}\right) = 1 - \Phi(0.04) = 0.4840.$$

单个正态总体均值的检验

设 X_1, X_2, \dots, X_n 是来自正态分布 $\mathcal{N}(\mu, \sigma^2)$ 的样本, 考虑如下三种关于 μ 的检验问题:

$$\text{I} \quad H_0 : \mu \leq \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0$$

$$\text{II} \quad H_0 : \mu \geq \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$$

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

其中, μ_0 是已知常数. 由于正态总体含两个参数, 总体方差 σ^2 已知与否对检验有影响. 下面我们分 σ 已知和未知两种情况讨论.

单个正态总体均值的检验

下面我们分 σ 已知和未知两种情况讨论.

- σ 已知: 由于 μ 的点估计是样本均值 \bar{x} , 根据正态分布的性质选择检验量

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

给定显著性水平 α 后可得拒绝域, 这种检验方法称为 **Z 检验法**.

- σ 未知: 由于 σ 的点估计是无偏样本方差 S , 根据正态分布的性质选择检验量

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

给定显著性水平 α 后可得拒绝域, 这种检验方法称为 **t 检验法**.

单个正态总体均值的检验

条件	H_0	H_1	检验统计量	拒绝域
σ 已知	$\mu \leq \mu_0$	$\mu > \mu_0$	$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$\{Z \geq \mu_\alpha\}$
	$\mu \geq \mu_0$	$\mu < \mu_0$		$\{Z \leq -\mu_\alpha\}$
	$\mu = \mu_0$	$\mu \neq \mu_0$		$\{ Z \geq \mu_{\alpha/2}\}$
σ 未知	$\mu \leq \mu_0$	$\mu > \mu_0$	$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$	$\{t \geq t_\alpha(n-1)\}$
	$\mu \geq \mu_0$	$\mu < \mu_0$		$\{t \leq -t_\alpha(n-1)\}$
	$\mu = \mu_0$	$\mu \neq \mu_0$		$\{ t \geq t_{\alpha/2}(n-1)\}$

假设检验：例 0.155

例 0.155 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. ($\mu_{0.025} = 1.96$)

解答：例 0.155

题目：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. ($\mu_{0.025} = 1.96$)

解答：

- 提出假设: $H_0 : \mu = 4.55$ vs $H_1 : \mu \neq 4.55$.
- 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.

假设检验：例 0.156

例 0.156 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. ($\mu_{0.05} = 1.645$)

解答：例 0.156

题目：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. ($\mu_{0.05} = 1.645$)

解答:

- 提出假设: $H_0 : \mu \geq 1000$ vs $H_1 : \mu < 1000$.
- 若 H_0 成立, 选择检验量

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

求得拒绝域为 $Z < -\mu_\alpha = -1.645$.

- 计算样本均值可知 $\bar{X} = 960$, 于是有

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = -2 < -1.645$$

由此可拒绝 H_0 , 说明这批灯泡不合格.

假设检验：例 0.157

例 0.157 某厂生产的铝材长度服从正态分布, 假设其均值为 240cm. 现从一批铝材中随机抽取 5 件产品, 测得其长度 (单位:cm) 为 239.7, 239.6, 239, 240, 239.2, 试判断这批铝材的长度是否满足设定要求?

解答：例 0.157

题目：某厂生产的铝材长度服从正态分布，假设其均值为 240cm. 现从一批铝材中随机抽取 5 件产品，测得其长度 (单位:cm) 为 239.7, 239.6, 239, 240, 239.2, 试判断这批铝材的长度是否满足设定要求？

解答：

- 提出假设: $H_0 : \mu = 240$ vs $H_1 : \mu \neq 240$.
- 若 H_0 成立, 选择检验量

$$t = \frac{\bar{x} - 240}{S/\sqrt{n}} \sim t(4),$$

求得拒绝域为 $|t| \geq t_{\alpha/2}(n-1)$.

- 若取 $\alpha = 0.05$, 则查表得 $t_{0.025}(4) = -t_{0.975}(4) = -2.7764$. 计算样本均值可知 $\bar{X} = 239.5$, $S = 0.4$, 故

$$t = \frac{239.5 - 240}{0.4/\sqrt{5}} = -2.795 < -2.7764$$

由此可拒绝 H_0 , 说明这批铝材的长度不满足设定要求.

两个正态总体均值差的检验

设 X_1, X_2, \dots, X_n 是来自正态分布 $\mathcal{N}(\mu_1, \sigma_1^2)$ 的样本, y_1, y_2, \dots, y_m 是来自正态分布 $\mathcal{N}(\mu_2, \sigma_2^2)$ 的样本, 两个样本相互独立. 考虑如下三种检验问题:

$$\text{I} \quad H_0 : \mu_1 - \mu_2 \leq 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 > 0$$

$$\text{II} \quad H_0 : \mu_1 - \mu_2 \geq 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 < 0$$

$$\text{III} \quad H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq 0$$

下面我们分 σ_1 和 σ_2 已知和相等但未知两种情况讨论.

两个正态总体均值差的检验

下面我们分 σ_1 和 σ_2 已知和相等但未知两种情况讨论.

- σ_1 和 σ_2 已知: 由于 $\mu_1 - \mu_2$ 的点估计 $\bar{x} - \bar{y}$ 的分布已知

$$\bar{x} - \bar{y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right),$$

根据正态分布的性质选择检验量

$$U = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1).$$

给定显著性水平 α 后可得拒绝域.

两个正态总体均值差的检验

- $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知: 有

$$\bar{x} - \bar{y} \sim \mathcal{N} \left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right), \quad \frac{(n-1)S_1^2 + (m-1)S_2^2}{\sigma^2} \sim \chi^2(m+n-2),$$

故可以选择检验量

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{S_W \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2),$$

其中

$$S_W = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2},$$

给定显著性水平 α 后可得拒绝域.

两个正态总体均值差的检验

条件	H_0	H_1	检验统计量	拒绝域
σ_1 和 σ_2 已知	$\mu_1 - \mu_2 \leq 0$	$\mu_1 - \mu_2 > 0$	$U = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$	$\{U \geq \mu_\alpha\}$
	$\mu_1 - \mu_2 \geq 0$	$\mu_1 - \mu_2 < 0$		$\{U \leq -\mu_\alpha\}$
	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq \mu_0$		$\{ U \geq \mu_{\alpha/2}\}$
$\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知	$\mu_1 - \mu_2 \leq 0$	$\mu_1 - \mu_2 > 0$	$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{S_W \sqrt{\frac{1}{n} + \frac{1}{m}}}$	$\{t \geq t_\alpha(m + n - 2)\}$
	$\mu_1 - \mu_2 \geq 0$	$\mu_1 - \mu_2 < 0$		$\{t \leq -t_\alpha(m + n - 2)\}$
	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq \mu_0$		$\{ t \geq t_{\alpha/2}(m + n - 2)\}$

假设检验：例 0.158

例 0.158 某厂铸造车间为提高零件的耐磨性, 试制了一种镍合金零件以取代铜合金零件, 为此从两种零件中各抽取容量分别为 8 和 9 的样本, 测得其耐磨性为下表. 假设两类零件的耐磨性服从正态分布, 且方差相等, 在显著性水平 $\alpha = 0.05$ 的情况下, 检验判断镍合金的耐磨性是否有明显提高.

镍合金	76.43	76.21	73.58	69.69	65.29	70.83	82.75	72.34	
铜合金	73.66	64.27	69.34	71.37	69.77	68.12	67.27	68.07	62.61

解答：例 0.158

解答：

- 用 X 表示镍合金的耐磨性, Y 表示铜合金的耐磨性, 且 $X \sim \mathcal{N}(\mu_1, \sigma^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma^2)$. 提出假设: $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$.
- 由于两者的方差相等但未知, 选择 t 检验量

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{S_W \sqrt{\frac{1}{n} + \frac{1}{m}}},$$

经计算有 $\bar{x} = 73.39$, $\bar{y} = 68.2756$, $\sum_{i=1}^8 (X_i - \bar{x})^2 = 191.7958$, $\sum_{i=1}^9 (y_i - \bar{y})^2 = 91.1548$.
从而, $S_W = \sqrt{\frac{1}{8+9-2}(191.7958 + 91.1548)} = 4.3432$,

$$t = \frac{73.39 - 68.2756}{4.3432 \times \sqrt{\frac{1}{8} + \frac{1}{9}}} = 2.4234.$$

- 查表可知 $t_{0.975}(15) = 1.7531 < 2.4234$, 故拒绝原假设, 即判断镍合金耐磨性有所提高.

成对数据检验

在很多实际应用中, 为了比较两个总体之间的差异, 往往会得到一批成对的数据, 然后基于观察的数据分析判断两个总体之间是否有显著的区别, 这种方法称为成对 (pairwise) 比较法.

假设观察到 n 对相互独立的随机变量 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, 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

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

并建立假设 $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$, 方差 σ^2 未知考虑 t 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: $|t| > t_{\alpha/2}(n-1)$.

假设检验：例 0.159

例 0.159 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. 问这两种方法在 $\alpha = 0.05$ 下是否有显著区别?

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

解答：例 0.159

题目：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. 问这两种方法在 $\alpha = 0.05$ 下是否有显著区别?

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

解答:

- 设 $Z_i = X_i - Y_i (i \in [10])$, 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.

单个正态总体方差的检验

设 X_1, X_2, \dots, X_n 是来自正态分布 $\mathcal{N}(\mu, \sigma^2)$ 的样本. 假定 μ 未知, 采用 χ^2 检验统计量

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

在显著性水平为 α 时求解拒绝域, 这种检验方法称为 χ^2 检验法.

- 假设 $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ 的拒绝域为

$$\{\chi^2 \geq \chi_{\alpha/2}^2(n-1)\} \cup \{\chi^2 \leq \chi_{1-\alpha/2}^2(n-1)\}$$

- 假设 $H_0 : \sigma^2 \geq \sigma_0^2$ vs $H_1 : \sigma^2 < \sigma_0^2$ 的拒绝域为

$$\{\chi^2 \leq \chi_{1-\alpha}^2(n-1)\}$$

- 假设 $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$ 的拒绝域为

$$\{\chi^2 \geq \chi_{\alpha}^2(n-1)\}$$

假设检验：例 0.160

例 0.160 某类钢板每块的重量 X 服从正态分布, 质量指标要求钢板重量的方差不得超过 0.016. 现从某批钢板中随机抽取 25 块, 得其样本方差 $S^2 = 0.025$, 问该批钢板的重量是否满足指标要求?

解答：例 0.160

题目：某类钢板每块的重量 X 服从正态分布, 质量指标要求钢板重量的方差不得超过 0.016. 现从某批钢板中随机抽取 25 块, 得其样本方差 $S^2 = 0.025$, 问该批钢板的重量是否满足指标要求?

解答:

- 提出假设: $H_0 : \sigma^2 \leq 0.016$ vs $H_1 : \sigma^2 > 0.016$.
- 查表得 $\chi_{0.95}^2(24) = 36.415$, 若 H_0 成立, 选择检验量

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{24 \times 0.025}{0.016} = 37.5 > 36.415.$$

由此可拒绝 H_0 , 说明批钢板的重量不满足要求.

两个正态总体方差比的检验

设 X_1, X_2, \dots, X_n 是来自正态分布 $\mathcal{N}(\mu_1, \sigma_1^2)$ 的样本, y_1, y_2, \dots, y_m 是来自正态分布 $\mathcal{N}(\mu_2, \sigma_2^2)$ 的样本. 假定 μ_1, μ_2 未知, 无偏样本方差分别为 S_x^2, S_y^2 , 由此可建立如下检验统计量:

$$F = \frac{S_x^2}{S_y^2} \sim F(n-1, m-1), \quad \text{当 } \sigma_1^2 = \sigma_2^2 \text{ 时}$$

在显著性水平为 α 时求解拒绝域:

- 假设 $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$ 的拒绝域为 $\{F \leq F_{\alpha/2}(n-1, m-1)\} \cup \{F \geq F_{1-\alpha/2}(n-1, m-1)\}$.
- 假设 $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs $H_1 : \sigma_1^2 < \sigma_2^2$ 的拒绝域为 $\{F \leq F_{\alpha}(n-1, m-1)\}$.
- 假设 $H_0 : \sigma_1^2 \leq \sigma_2^2$ vs $H_1 : \sigma_1^2 > \sigma_2^2$ 的拒绝域为 $\{F \geq F_{1-\alpha}(n-1, m-1)\}$.

假设检验：例 0.161

例 0.161 甲乙两台机床加工某种零件, 零件的直径服从正态分布, 总体方差反映了加工精度, 为比较两台机床的加工精度有无差别, 现从各自加工的零件中分别抽取 7 件产品和 8 件产品, 测得其直径为

甲机床 16.2 16.8 15.8 15.5 16.7 15.6 15.8

乙机床 15.9 16 16.4 16.1 16.5 15.8 15.7 15

解答：例 0.161

题目：甲乙两台机床加工某种零件，零件的直径服从正态分布，总体方差反映了加工精度，为比较两台机床的加工精度有无差别，现从各自加工的零件中分别抽取 7 件产品和 8 件产品，测得其直径为

甲机床	16.2	16.8	15.8	15.5	16.7	15.6	15.8	
乙机床	15.9	16	16.4	16.1	16.5	15.8	15.7	15

解答：

- 提出假设： $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$.
- 经计算有 $S_x^2 = 0.2729$, $S_y^2 = 0.2164$, 于是有 $F = 0.2729/0.2164 = 1.261$. 查表得 $F_{0.975}(6, 7) = 5.12$, $F_{0.025}(6, 7) = 1/F_{0.975}(7, 6) = 0.175$. 其拒绝域为

$$W = \{F \leq 0.175\} \cup \{F \geq 5.12\}$$

样本未落入拒绝域，可认为两台机床的加工精度无差别.

假设检验与置信区间的关系

不同点: 假设检验与置信区间最大区别在于解决问题的不相同:

- 假设检验: 根据样本信息判断关于总体的某个假设是否正确.
- 置信区间: 估计未知参数的取值范围.

相似点: 检验统计量与枢轴量的构造相似, 显著性水平 α 的假设检验不拒绝域的边界与置信水平为 $1 - \alpha$ 的区间估计的上下限有对应关系.

- 对于 σ 未知时单个正态总体均值的双边检验, 显著性水平 α 对应的拒绝域为 $\{|t| \geq t_{\alpha/2}(n-1)\}$, 因此不拒绝域可写为

$$\overline{W} = \left\{ \bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) < \mu_0 < \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) \right\}$$

- 给定置信度 $1 - \alpha$, σ 未知时 μ 的置信区间为

$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) \right]$$

非参假设检验

前面介绍的各种检验都是在总体服从正态分布前提下, 对参数进行假设检验. 实际中可能遇到这样的情形: 总体服从何种理论分布并不知道, 要求我们直接对总体分布提出一个假设.

设 X_1, X_2, \dots, X_n 是来自总体 $F(x)$ 的样本, 检验原假设

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

其中 $F_0(x)$ 是形式已知但含有若干个未知参数的分布函数. 这个分布检验问题就是检验观测数据是否与理论分布相符. 这一类检验问题统称为分布的拟合检验, 它们是一类非参数检验问题.

- 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, 应先用最大似然估计/矩估计估计参数.

分布的 χ^2 拟合优度检验

当样本容量较大时, 分布的拟合检验可以用 χ^2 拟合优度检验来解决. 下面介绍 χ^2 拟合优度检验法:

将随机试验结果全体 Ω 分为 k 个互不相容的事件 A_1, A_2, \dots, A_k , 并使得落入每个 A_i 的样本个数不小于 5, 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, X_2, \dots, X_n , 事件 A_i 出现的频率为 n_i/n . 当 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, 皮尔逊构造了检验量:

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

称为皮尔逊 χ^2 统计量.

分布的 χ^2 拟合优度检验

定理 0.77 若分布 $F_0(x)$ 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

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

给定显著性水平 α , 若 $W > \chi_\alpha^2(k-1)$ 则拒绝 H_0 .

定理 0.78 当 $n \rightarrow +\infty$ 时, 有 $W \xrightarrow{d} \chi^2(k-r-1)$ 成立.

假设检验：例 0.162

例 0.162 某试验有四种不同的结果 $\{A, B, C, D\}$. 现进行如下实验：独立重复实验直到结果 A 发生为止. 试验 200 次, 记录抛掷的次数结果如下表, 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

解答：例 0.162

题目：如上所述.

解答：

- 提出假设: 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

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

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

- 计算统计检验量

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

根据统计量实值 $W > \chi_{0.05}^2(4) = 9.488$ 则拒绝 H_0 , 该试验不服从均匀分布.

假设检验：例 0.163

例 0.163 1911 年著名物理学家卢瑟福等人为探索原子的内部结构进行了一项实验,即用一束带正电的、质量比电子大得多的高速运动的 α 粒子轰击金箔,证明了正电荷集中在原子中心. 考察下表中卢瑟福实验的数据,是以 7.5s 为时间单位所做的 2608 次观察所得的数据,观测的是一枚放射性 α 物质在单位时间内放射的质点数. 试问 7.5s 中放射出的 α 质点数是否服从泊松分布 $P(\lambda)$?

质点数 k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
观察数 n_k	57	203	383	525	532	408	273	139	45	27	10	4	2	0	0

解答：例 0.163

解答：

- 提出假设: H_0 : 7.5s 中放射出的 α 质点数服从泊松分布 $P(\lambda)$, 又泊松分布参数 λ 的最大似然估计为 $\hat{\lambda} = \bar{x}$, 即

$$\hat{\lambda} = \frac{1}{n} \sum_{k=1}^n X_i = \frac{\sum_{k=0}^{14} kn_k}{\sum_{k=0}^{14} n_k} = 3.87$$

- 计算泊松分布的概率估计值

$$\hat{p}_k = \frac{\hat{\lambda}^k}{k!} e^{-\hat{\lambda}}, \quad k = 1, 2, \dots$$

为了满足每一类出现的样本观测次数不小于 5, 我们把 $k \geq 11$ 作为一类, 记为第 12 类, 可以得到检验统计量的值为

$$W = \sum_{i=1}^{12} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = 12.8967.$$

此处分布自由度为 $12 - 1 - 1 = 10$, 对 $\alpha = 0.05$, 查表得 $W \leq \chi_{0.95}^2(10) = 18.3070$,

因此不能拒绝原假设 H_0 , 可以认为该放射物质在 7.5 秒时间内放射的 α 质点数与泊松分布吻合.

• 附: 分布拟合检验计算过程:

序号	质点数	观测数 n_i	概率估计 \hat{p}_i	$(n_i - n\hat{p}_i)^2/n\hat{p}_i$
1	0	57	0.0209	0.1147
2	1	203	0.0807	0.2672
3	2	383	0.1562	1.4614
4	3	525	0.2015	0.0005
5	4	532	0.195	1.0766
...
11	10	10	0.0043	0.1286
12	≥ 11	6	0.0022	0.0158

假设检验：例 0.164

例 0.164 某工厂生产一种滚珠, 现随机抽取 50 件产品, 测得其直径 (单位:mm) 为

15.0	15.8	15.2	15.1	15.9	14.7	14.8	15.5	15.6	15.3
15.0	15.6	15.7	15.8	14.5	15.1	15.3	14.9	14.9	15.2
15.9	15.0	15.3	15.6	15.1	14.9	14.2	14.6	15.8	15.2
15.2	15.0	14.9	14.8	15.1	15.5	15.5	15.1	15.1	15.0
15.3	14.7	14.5	15.5	15.0	14.6	14.6	14.2	14.2	14.5

问滚珠直径是否服从正态分布?

解答：例 0.164

解答：

- 设滚珠直径为 X , 提出假设: $H_0 : F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$, 由样本数据求得 μ, σ 的最大似然估计为 $\hat{\mu} = 15.1, \hat{\sigma}^2 = 0.4379^2$.
- 根据样本数据特点并考虑到各组观测值个数不低于 5, 去分点为

$$a_0 = -\infty, \quad a_1 = 14.55, \quad a_2 = 14.95, \quad a_3 = 15.35, \quad a_4 = 15.75, \quad a_5 = +\infty$$

由此把数据分为 5 组, 各组数据个数分别为

$$n_1 = 6, \quad n_2 = 11, \quad n_3 = 20, \quad n_4 = 8, \quad n_5 = 5$$

- 利用公式

$$\hat{p}_i = \Phi\left(\frac{a_i - 15.1}{0.4379}\right) - \Phi\left(\frac{a_{i-1} - 15.1}{0.4379}\right), \quad i = 1, 2, 3, 4, 5$$

求得

$$\hat{p}_1 = 0.104559, \quad \hat{p}_2 = 0.261412, \quad \hat{p}_3 = 0.349998, \quad \hat{p}_4 = 0.215174, \quad \hat{p}_5 = 0.068857$$

- 可以得到检验统计量的值为

$$W = \sum_{i=1}^5 \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = 2.2109.$$

此处分布自由度为 $5 - 2 - 1 = 2$, 对 $\alpha = 0.05$, 查表得 $\chi_{0.95}^2(2) = 5.9915 > 2.2109$, 因此不能拒绝原假设 H_0 .

列联表的独立性检验

下面我们分析按两个或多个特征分类的频数数据, 这种数据通常称为交叉分类数据, 它们一般以表格形式给出, 称为列联表. 列联表分析的基本问题是, 考察各属性之间有无关联, 即判别两属性是否独立. 例如: 色盲与其性别是否有关?

若考虑的属性多于两个, 称为多维列联表. 本次我们只讨论二维列联表, 列联表分析在应用统计, 特别在医学, 生物学及社会科学中有广泛的应用.

列联表的独立性检验

一般的, 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 与 Y 的取值分为 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 $n_{i.} = \sum_{j=1}^s n_{ij}$ 和 $n_{.j} = \sum_{i=1}^r n_{ij}$ 为行和或列和, 则 $n = \sum_{i,j} n_{ij}$. 建立如下二维列联表:

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

列联表的独立性检验

下面介绍列联表的独立性检验法: 提出假设 H_0 : X 与 Y 相互独立. 记

$$p_{ij} = \Pr(X \in A_i, Y \in B_j)$$

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

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

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

设计假设检验统计量

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

在显著性水平 α 时有 $W \sim \chi^2((r-1)(s-1))$ 成立, 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 相互独立.

假设检验：例 0.165

例 0.165 为研究儿童智力发展与营养得关系, 某研究机构调查了 1436 名儿童, 得到如下数据, 试在显著性水平 0.05 下判断智力发展与营养有无关系.

	智商				合计
	< 80	80 ~ 89	90 ~ 99	≥ 100	
营养良好	367	342	266	329	1304
营养不良	56	40	20	16	132
合计	423	382	286	345	1436

解答：例 0.165

解答：

- 用 A 表示营养状况, 它有两个水平: A_1 表示营养良好, A_2 表示营养不良; B 表示儿童智商, 它有四个水平: B_1, B_2, B_3, B_4 分别表示表中四种情况. 建立假设 H_0 : 营养状况与智商无关联, 即 A 与 B 是独立的.

$$H_0 : p_{ij} = p_{i.}p_{.j}, \quad i = 1, 2, j = 1, 2, 3, 4$$

在原假设 H_0 成立下, 我们可以计算诸参数得最大似然估计值

$$\hat{p}_{1.} = 1304/1436 = 0.9081, \quad \hat{p}_{2.} = 132/1436 = 0.0919$$

$$\hat{p}_{.1} = 432/1436 = 0.2946, \quad \hat{p}_{.2} = 382/1436 = 0.266$$

$$\hat{p}_{.3} = 286/1436 = 0.1992, \quad \hat{p}_{.4} = 345/1436 = 0.2403$$

- 进一步算出 $W = \sum_{i=1}^2 \sum_{j=1}^4 \frac{(n_{ij} - n\hat{p}_{i.}\hat{p}_{.j})^2}{n\hat{p}_{i.}\hat{p}_{.j}} = 19.2785$ 此处分布自由度为 $(2-1)(4-1) = 3$, 对 $\alpha = 0.05$, 查表得 $\chi_{0.95}^2(3) = 7.8147 < 19.2785$, 因此拒绝原假设 H_0 .