假设检验（8.4–8.7 节选讲）

李孟棠，朱彬

中山大学智能工程学院

limt29@mail.sysu.edu.cn, zhub26@mail.sysu.edu.cn

2021 年 x 月 x 日

主要内容

1. 置信区间和假设检验之间的关系
2. 样本容量的选取
3. 分布拟合检验
4. 假设检验问题的 p 值法

主要内容

置信区间和假设检验之间有明显的联系。先考察双侧置信区间和双边检验之间的对应关系。设 $X_1, \ldots, X_n$ 是一个来自总体的样本，$x_1, \ldots, x_n$ 是相应的观察值，集合 Θ 是参数 θ 的取值范围。

设 $(\underline{\theta}, \overline{\theta})$ 是参数 θ 的一个置信水平为 $1-\alpha \in (0,1)$ 的置信区间。则对于任意 $\theta \in \Theta$，有

$$P_\theta\left\{\underline{\theta}(X_1, \ldots, X_n) < \theta < \overline{\theta}(X_1, \ldots, X_n)\right\} \geq 1-\alpha. \tag{1}$$

考虑显著性水平为 $\alpha \in (0,1)$ 的双边检验

$$H_0: \ \theta = \theta_0, \quad H_1: \ \theta \neq \theta_0.$$

则由(1)知

$$P_{\theta_0}\left\{\underline{\theta}(X_1, \ldots, X_n) < \theta_0 < \overline{\theta}(X_1, \ldots, X_n)\right\} \geq 1-\alpha,$$

即有

$$P_{\theta_0}\left\{\{\theta_0 \le \underline{\theta}(X_1,\ldots,X_n)\} \cup \left\{\theta_0 \ge \overline{\theta}(X_1,\ldots,X_n)\right\}\right\} \le \alpha.$$

按照显著性水平为 α 的拒绝域定义，上述双边检验的拒绝域为

$$\theta_0 \le \underline{\theta}(x_1,\ldots,x_n) \quad \text{或} \quad \theta_0 \ge \overline{\theta}(x_1,\ldots,x_n).$$

也就是说，当我们要处理双边假设时，可以先求出 θ 的置信水平为 $1-\alpha$ 的置信区间 $(\underline{\theta},\overline{\theta})$，然后考察区间 $(\underline{\theta},\overline{\theta})$ 是否包含 θ_0。如果 $\theta_0 \in (\underline{\theta},\overline{\theta})$，则接受 H_0，否则拒绝 H_0。

反之，给定双边假设检验问题的解，我们可以得到未知参数的置信区间。类似地，单侧置信区间和单边检验问题的解也有这样的对应关系。详见书本 P. 196（作为延伸阅读）。

例

设 $X \sim N(\mu, 1)$，μ 未知。取 $\alpha = 0.05$。由一个容量 $n = 16$ 的样本算得 $\bar{x} = 5.2$，于是得到参数 μ 的一个置信水平为 0.95 的置信区间

$$\left(\bar{x} - \frac{1}{\sqrt{n}} z_{\alpha/2}, \bar{x} + \frac{1}{\sqrt{n}} z_{\alpha/2}\right) = (4.71, 5.69).$$

现在重新考虑假设检验问题

$$H_0: \ \mu = \mu_0 = 5.5, \quad H_1: \ \mu \neq \mu_0.$$

由于 $5.5 \in (4.71, 5.69)$，故接受 H_0。

例

数据如上例，即总体 $X \sim N(\mu, 1)$，μ 未知。由一个容量 $n = 16$ 的样本算得 $\bar{x} = 5.2$。试求右边检验问题

$$H_0: \ \mu \le \mu_0, \quad H_1: \ \mu > \mu_0$$

的接受域，并求 μ 的单侧置信下限（$\alpha = 0.05$）。

解：检验问题的拒绝域为

$$z = \frac{\bar{x} - \mu_0}{1/\sqrt{n}} \ge z_\alpha,$$

即 $\mu_0 \le 4.79$。于是检验问题的接受域为 $\mu_0 > 4.79$。如此得到参数 μ 的单侧置信区间 $(4.79, \infty)$ 和单侧置信下限 $\underline{\mu} = 4.79$。

主要内容

到目前为止，我们在进行假设检验时总是根据问题的要求，预先给出显著性水平 α 以控制犯第 I 类错误的概率。

在一些实际问题中，我们除了希望控制犯第 I 类错误的概率外，往往还希望控制犯第 II 类错误的概率。

本节阐明如何选取合适的样本容量使得犯第 II 类错误的概率被控制在预先给定的限度以内。

Definition (操作特征 (OC) 函数)

若对参数 θ 的某假设检验问题采用检验法 C，那么函数

$$\beta(\theta) := P_\theta\{\text{接受}H_0\}$$

称为检验法 C 的操作特征（operating characteristic，简称 OC）函数，其图形称为 OC 曲线。

上述命名来自始于 1941 年的军用雷达操作。

由定义知，如果检验法 C 的显著性水平为 α，那么

- 当真值 $\theta \in H_0$ 时，$\beta(\theta)$ 就是做出正确判断（即 H_0 为真时接受 H_0）的概率，故此时 $\beta(\theta) \geq 1-\alpha$；
- 当 $\theta \in H_1$ 时，$\beta(\theta)$ 就是犯第 II 类错误的概率，而 $1-\beta(\theta)$ 是做出正确判断（即 H_0 为假时拒绝 H_0）的概率。

定义（功效函数）

我们称 $1-\beta(\theta)$ 为检验法 C 的**功效（power）函数**。易知

$$1-\beta(\theta) := P_\theta\{\text{拒绝}H_0\text{，即样本观察值落在拒绝域内}\}$$

当 $\theta^* \in H_1$ 时，函数值 $1-\beta(\theta^*)$ 称为检验法 C 在点 θ^* 的**功效**(power of the test)，也就是当真实参数 $\theta=\theta^*$ 时，检验法 C 做出正确判断的概率。

这里只给出正态总体均值检验法的 OC 函数。

1. Z 检验法的 OC 函数

先考虑右边检验问题 $H_0: \mu \le \mu_0$，$H_1: \mu > \mu_0$，其 OC 函数为

$$\beta(\mu) = P_\mu\{\text{接受}H_0\} = P_\mu\left\{\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < z_\alpha\right\}$$

$$= P_\mu\left\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_\alpha - \underbrace{\frac{\mu-\mu_0}{\sigma/\sqrt{n}}}_{=:\lambda}\right\} = \Phi(z_\alpha - \lambda),$$

其中 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$。函数图形见书本 P. 198（注意：根据定义 $\beta(0) < 1$，书上在这一点上画错了）。

此 OC 函数具有如下性质：

1. 它是 $\lambda = \frac{\mu-\mu_0}{\sigma/\sqrt{n}}$（也是 μ）的单调递减（连续）函数。
2. $\lim_{\mu\to\mu_0}\beta(\mu) = 1-\alpha$，$\lim_{\mu\to\infty}\beta(\mu) = 0$。

由函数 $\beta(\mu)$ 的连续性可知，当参数真值 μ 在 $\mu > \mu_0$ 区域（对应于备选假设 H_1）趋向于 μ_0 时，$\beta(\mu) \approx 1-\alpha$，检验法的功效为 $1-\beta(\mu) \approx \alpha$；由于 α 通常取得很小，这表明此时犯第 II 类错误的概率很大。

需要明确的是：无论样本容量 n 多大，不可能对所有参数 $\mu \in H_1$ 都把犯第 II 类错误的概率控制得很小。

可以做到：利用 OC 函数 $\beta(\mu)$ 确定合适的样本容量 n，使得当真值 $\mu \geq \mu_0 + \delta$（其中 $\delta > 0$ 为取定的值）时，犯第 II 类错误的概率不超过给定的 $\beta_0 \in (0,1)$，例如 0.05.

由于 $\beta(\mu)$ 是 μ 的递减函数，所以当 $\mu \geq \mu_0 + \delta$ 时有

$$\beta(\mu_0 + \delta) \geq \beta(\mu).$$

于是只需要控制 $\beta(\mu_0 + \delta) = \Phi(z_\alpha - \sqrt{n}\delta/\sigma) \leq \beta_0$，也就是

$$z_\alpha - \sqrt{n}\,\delta/\sigma \leq -z_{\beta_0},$$

即

$$\sqrt{n} \geq \frac{(z_\alpha + z_{\beta_0})\sigma}{\delta}.$$

此时，对于任意的参数 $(H_1 \ni)\mu \geq \mu_0 + \delta$，$Z$ 检验法犯第 II 类错误的概率不超过事先指定的 β_0。

类似可得左边检验问题 $H_0: \mu \geq \mu_0$，$H_1: \mu < \mu_0$ 的 OC 函数为

$$\beta(\mu) = \Phi(z_\alpha + \lambda), \quad \text{其中}\lambda = \frac{\mu - \mu_0}{\sigma/\sqrt{n}}.$$

当参数真值 $\mu < \mu_0$ 时，$\beta(\mu)$ 给出犯第 II 类错误的概率。

给定 $\delta > 0$，当样本容量 n 满足

$$\sqrt{n} \geq \frac{(z_\alpha + z_{\beta_0})\sigma}{\delta} \quad \text{（与前面相同！）}$$

时，对于任意的参数 $(H_1 \ni)\mu \leq \mu_0 - \delta$，$Z$ 检验法犯第 II 类错误的概率不超过事先指定的 β_0。

Ex. 验证以上结论。

关于双边检验问题 $H_0: \mu = \mu_0$，$H_1: \mu \neq \mu_0$ 的分析见书本 P. 199。

例 (工业产品质量抽检方案)

设有一大批产品，产品质量指标 $X \sim N(\mu, \sigma^2)$，以 μ 小者为佳。厂方要求所确定的验收方案对高质量的产品 $(\mu \leq \mu_0)$ 能以高概率 $1-\alpha$ 为买方所接受，买方则要求低质产品 $(\mu \geq \mu_0 + \delta,\ \delta > 0)$ 能以高概率 $1-\beta$ 被拒绝，其中 α, β 由厂方与买方协商给出。拟采取一次抽样以确定该批产品是否为买方所接受，问应该如何安排抽样方案（确定检验法和样本容量）？已知 $\mu_0 = 120$，$\delta = 20$，且由工厂长期经验知 $\sigma^2 = 900$，又经商定 α, β 均取为 0.05。

解：题中检验问题即为右边检验

$$H_0:\ \mu \leq \mu_0,\ H_1:\ \mu > \mu_0,$$

且要求

- 当 $\mu \leq \mu_0$ 时能以 $1-\alpha = 0.95$ 的概率接受 H_0；
- 当 $\mu \geq \mu_0 + \delta$ 时能以 $1-\beta = 0.95$ 的概率拒绝 H_0。

本题总体方差已知，故采用 Z 检验法。由前面关于右边检验的分析可知，给定 α，β，δ 时样本容量应满足

$$\sqrt{n} \geq \frac{(z_\alpha + z_\beta)\sigma}{\delta} = \frac{(2 \times 1.645) \times 30}{20} = 4.935,$$

所以取 $n = 25$。零假设的拒绝域形式为

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha = 1.645,$$

即当样本均值 $\bar{x} \geq 129.87$ 时，买方就拒绝这批产品；反之就接受这批产品。

2. t 检验法的 OC 函数

当总体方差未知时，对于右边检验问题 $H_0: \mu \leq \mu_0$，$H_1: \mu > \mu_0$ 使用 t 检验法，其 OC 函数为

$$\beta(\mu) = P_\mu\{\text{接受}H_0\} = P_\mu\left\{\frac{\bar{X}-\mu_0}{S/\sqrt{n}} < t_\alpha(n-1)\right\}. \tag{2}$$

再次引入参数 $\lambda = \frac{\mu-\mu_0}{\sigma/\sqrt{n}}$，则(2)中随机变量

$$\frac{\bar{X}-\mu_0}{S/\sqrt{n}} = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} + \lambda\right) \Big/ \left(\frac{S}{\sigma}\right)$$

服从非中心参数为 λ、自由度为 $n-1$ 的非中心 t 分布 (C. C. Craig, 1941)。在 $\lambda = 0$ 时，它退化为通常的 $t(n-1)$ 变量。

注意：不同于 Z 检验法，t 检验法的 OC 函数没有简单的解析形式，而需要近似计算。对于常用的样本容量 n，已有制作好的 OC 曲线和表格可用。

对于给定的 α, β 和 $\delta > 0$，可查询书末附表 7 得到所需的样本容量 n，使得:

- 对于右边检验问题，当 $\mu \in H_1$ 且 $\frac{\mu-\mu_0}{\sigma} \geq \delta$ 时，t 检验法犯第 II 类错误的概率不超过 β；
- 对于左边检验问题，当 $\mu \in H_1$ 且 $\frac{\mu-\mu_0}{\sigma} \leq -\delta$ 时，t 检验法犯第 II 类错误的概率不超过 β；
- 对于双边检验问题，当 $\mu \in H_1$ 且 $\frac{|\mu-\mu_0|}{\sigma} \geq \delta$ 时，t 检验法犯第 II 类错误的概率不超过 β；

附表 7 的结构表明：对于给定的 β 和 $\delta > 0$，

- 相同显著性水平 α 的右边、左边检验所需的样本容量相同；
- 显著性水平为 α 的单边检验和显著性水平为 2α 的双边检验所需的样本容量（近似）相同。

延伸阅读:

- 书本 P. 200–201 例 2、例 3;
- 样本容量的确定依赖于参数 δ。例如在右边检验中，通常给定 $\mu_1 > \mu_0$，再计算 $\delta = \frac{\mu_1 - \mu_0}{\sigma}$。然而在 t 检验法处理的问题中方差 σ^2 未知，因此在仅给定 μ_1 时无法计算 δ。此时我们可以采用一种近似方法，见书本 P. 201。
- 两个正态总体均值差的 t 检验法中样本容量的确定和 P. 201–202 例 4。

主要内容

前面介绍的各种检验法都建立在总体分布形式已知的前提下。但在实际问题中，有时不知道总体分布的类型，而需要根据样本来检验关于总体分布的假设。

本节介绍 χ^2 拟合检验法。它可以用来检验总体是否具有某一个指定的分布或属于某一个分布族。

单个分布的 χ^2 拟合检验法

设总体 X 的分布未知，$x_1, \ldots, x_n$ 是来自 X 的样本值。检验如下假设

$$H_0：总体\ X\ 的分布函数为\ F(x),$$
$$H_1：总体\ X\ 的分布函数不是\ F(x),$$

其中 $F(x)$ 不含未知参数，也可以用分布律或概率密度代替 $F(x)$；备选假设 H_1 可以不明确给出。

下面来定义检验统计量。

1 在假设 H_0 下，把 X 可能取值的全体 $\Omega \subset \mathbb{R}$ 分成互不相交的子集 $A_1, \ldots, A_k$，以 n_i $(i = 1, \ldots, k)$ 记样本观察值 $x_1, \ldots, x_n$ 落在 A_i 中的个数。

2 定义事件 $E_i = \{X$ 的值落在集合 A_i 内$\}$，则 n_i 为事件 E_i 在 n 次独立试验中发生的次数，发生的频率为 n_i/n。（显然 $\sum_{i=1}^{k} n_i = n$。）

3 当 H_0 为真时，我们可以根据分布函数 $F(x)$ 来计算事件 E_i 的概率，例如

$$p_i := P(E_i) = \int_{A_i} f(x)\mathrm{d}x, \quad i = 1, \dots, k,$$

其中 $f(x)$ 为 X 的概率密度函数。

4 根据大数定律，如果 H_0 为真且样本容量充分大，那么频率 n_i/n 和概率 p_i 应该比较接近，也就是 $\left(\frac{n_i}{n} - p_i\right)^2$ 不应太大。因此我们采用统计量

$$\sum_{i=1}^{k} C_i \left(\frac{n_i}{n} - p_i\right)^2$$

来度量样本与零假设中分布的吻合程度，其中 $C_i > 0$ 为待定常数。

Karl Pearson (1900) 证明了如下结果

定理 (Pearson)

取常数 $C_i = n/p_i$ $(i = 1, \ldots, k)$，即采用统计量

$$\chi^2 := \sum_{i=1}^{k} \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2 = \sum_{i=1}^{k} \frac{n_i^2}{np_i} - n.$$

如果样本容量 n 充分大 (≥ 50)，那么当 H_0 为真时上述统计量近似服从 $\chi^2(k-1)$ 分布。

证明略。注意此处统计量中的随机量为 n_i。

根据前面的讨论，当 H_0 为真时统计量 χ^2 的观察值不应太大；如果 χ^2 过大就拒绝 H_0，故拒绝域的形式为

$$\chi^2 \geq G \quad (G > 0 \text{为待定常数}).$$

对于给定的显著性水平 α，确定常数 G 使得

$$P_{H_0}\{\text{拒绝}H_0\} = P_{H_0}\{\chi^2 \geq G\} = \alpha.$$

由 Pearson 定理知 $G = \chi^2_\alpha(k-1)$。也就是说，如果样本观察值使得

$$\chi^2 \geq \chi^2_\alpha(k-1),$$

就在显著性水平 α 下拒绝 H_0，否则就接受 H_0。

称上述方法为单个分布的 χ^2**拟合检验法**。

使用注意点：

- 此检验法基于 Pearson 定理，因此必须保证样本容量 $n \geq 50$。
- 应保证 $np_i \geq 5$，否则应适当合并 A_i，以满足此条件（见例子）。

例

下表列出了某地区在夏季的一个月中由 100 个气象站报告的雷暴雨次数。

i	0	1	2	3	4	5	≥ 6
n_i	22	37	20	13	6	2	0
A_i	A_0	A_1	A_2	A_3	A_4	A_5	A_6

其中 n_i 是报告雷暴雨次数为 i 的气象站数。试用 χ^2 拟合检验法检验雷暴雨的次数（一个随机变量）X 是否服从均值 $\lambda = 1$ 的泊松分布（取显著性水平 $\alpha = 0.05$）。

解：按题意需检验假设

$$H_0: \ P\{X = i\} = \frac{\lambda^i e^{-\lambda}}{i!} = \frac{e^{-1}}{i!} \quad i = 0, 1, \cdots .$$

样本容量 $n = 100$。在 H_0 下 X 的所有可能取值为 $\Omega = \{0, 1, \cdots\}$，后者被分为两两互不相交的子集 $A_0, A_1, \ldots, A_6$。计算如下概率：

$$p_i := P\{E_i\} = P\{X = i\} = \frac{e^{-1}}{i!}, \quad i = 0, 1, \ldots, 5.$$

例如

$$p_0 = P\{X = 0\} = e^{-1} \approx 0.36788,$$

$$p_3 = P\{X = 3\} = \frac{e^{-1}}{3!} \approx 0.06131,$$

以及

$$p_6 = P\{X \geq 6\} = 1 - \sum_{i=0}^{5} p_i = 0.00059.$$

关于 χ^2 检验统计量的计算结果见下一页表格，其中有些 $np_i < 5$ 的组予以适当合并，使得每组均有 $np_i \geq 5$，如表中第 4 列花括号所示。

A_i	n_i	p_i	np_i		$n_i^2/(np_i)$
$A_0: \{X=0\}$	22	e^{-1}	36.788		13.16
$A_1: \{X=1\}$	37	e^{-1}	36.788		37.21
$A_2: \{X=2\}$	20	$e^{-1}/2$	18.394		21.75
$A_3: \{X=3\}$	13	$e^{-1}/6$	6.131	8.03	54.92
$A_4: \{X=4\}$	6	$e^{-1}/24$	1.533		
$A_5: \{X=5\}$	2	$e^{-1}/120$	0.307		
$A_6: \{X\geq 6\}$	0	$1-\sum_{i=0}^{5} p_i$	0.059		
					$\sum = 127.04$

并组后 $k=4$，χ^2 的自由度为 $k-1=3$，查表得 $\chi^2_{0.05}(k-1)=7.815$。现有 $\chi^2=\sum_{i=1}^{k}\frac{n_i^2}{np_i}-n=27.04>7.815$，即样本观察值在拒绝域内。因此在显著性水平 $\alpha=0.05$ 下拒绝 H_0，即认为样本不是来自均值 $\lambda=1$ 的泊松分布。

延伸阅读:

- P. 204–205 例 2
- 分布族的 χ^2 拟合检验（书本 P. 205–210）

在前面讨论的分布拟合检验中，我们假设待检验的分布函数 $F(x)$ 已知，这种情况在实际问题中不多见。我们经常遇到的原假设具有如下形式

$$H_0: \text{ 总体}X\text{的分布函数是}F(x;\theta_1,\ldots,\theta_r),$$

其中 F 的形式已知，而 $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_r)\in\Theta\subset\mathbb{R}^r$ 是未知参数（向量）。因此 $F(x;\boldsymbol{\theta})$ 实际上代表一族以 $\boldsymbol{\theta}$ 为参数的分布函数

$$\{F(x;\boldsymbol{\theta}):\boldsymbol{\theta}\in\Theta\}.$$

上面的假设 H_0 表示总体 X 的分布属于该分布族。可采用类似前面的 χ^2 检验法来处理此类问题。

主要内容

以上讨论的假设检验方法称为临界值法。本节介绍另一种检验方法，称为p 值法。

例

设总体 $X \sim N(\mu, \sigma^2)$，其中参数 μ 未知，$\sigma^2 = 100$。现有样本 $x_1, x_2, \ldots, x_{52}$，算得 $\bar{x} = 62.75$。要求检验假设

$$H_0: \ \mu \leq \mu_0 = 60, \quad H_1: \ \mu > \mu_0.$$

解：采用 Z 检验法，检验统计量为

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

代入数据，算得 Z 的观察值为 $z_0 = \frac{62.75-60}{10/\sqrt{52}} \approx 1.983$。计算概率

$$P_{H_0}\{Z \geq z_0\} \leq P_{H_0}\left\{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq z_0\right\} = 1 - \Phi(z_0) = 0.0238,$$

即为标准正态曲线下位于 z_0 右边的尾部面积。上式中随机变量 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，不等号来自于零假设中 $\mu \leq \mu_0$。

此概率称为 Z 检验法用于右边检验的 p 值，记为

$$p\text{值} = P_{H_0}\{Z \geq z_0\} = 0.0283.$$

- 若显著性水平 $\alpha \geq p$，则对应的临界值 $z_\alpha \leq z_0$，这表示观察值 z_0 落在拒绝域内，此时我们拒绝 H_0；
- 反之，若显著性水平 $\alpha < p$，则对应的临界值 $z_\alpha > z_0$，这表示观察值 z_0 不落在拒绝域内，此时我们接受 H_0。

上述分析表明，p 值 $= P_{H_0}\{Z \geq z_0\}$ 是原假设 H_0 可被拒绝的最小显著性水平。

Definition

假设检验问题的 p 值 (p-value) 是由检验统计量的样本观察值得出的原假设可被拒绝的最小显著性水平。

常见的假设检验问题的 p 值可以根据检验统计量的样本观察值和它在 H_0 下一个特定的参数值（一般是 H_0 与 H_1 所规定的参数的分界点）对应的分布求出。

例如在正态总体 $N(\mu,\sigma^2)$ 均值的检验中，当 σ^2 未知时，采用检验统计量

$$T=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}.$$

在右边、左边、双边检验问题中，当 $\mu=\mu_0$ 时 $T\sim t(n-1)$。如果由样本算得统计量 T 的观察值为 t_0，那么在检验问题

- $H_0:\ \mu\leq\mu_0$，$H_1:\ \mu>\mu_0$ 中，
  p 值 $=P_{\mu_0}\{T\geq t_0\}=t_0$ 右侧尾部面积；

- $H_0:\ \mu\geq\mu_0$，$H_1:\ \mu<\mu_0$ 中，
  p 值 $=P_{\mu_0}\{T\leq t_0\}=t_0$ 左侧尾部面积；

- $H_0:\ \mu=\mu_0$，$H_1:\ \mu\neq\mu_0$ 中，

  (i) 当 $t_0>0$ 时，

$$
\begin{aligned}
p\text{值} &= P_{\mu_0}\{|T|\geq t_0\} = P_{\mu_0}\{\{T\leq -t_0\}\cup\{T\geq t_0\}\} \\
&= 2\times(t_0\text{右侧尾部面积});
\end{aligned}
$$

  (ii) 当 $t_0<0$ 时，

$$
\begin{aligned}
p\text{值} &= P_{\mu_0}\{|T|\geq -t_0\} = P_{\mu_0}\{\{T\leq t_0\}\cup\{T\geq -t_0\}\} \\
&= 2\times(t_0\text{左侧尾部面积}).
\end{aligned}
$$

在现代计算机统计软件中，一般都会给出检验问题的 p 值。

按 p 值的定义，对于任意指定的显著性水平 α，有

- 若 p 值 $\leq\alpha$，则在显著性水平 α 下拒绝 H_0；
- 若 p 值 $>\alpha$，则在显著性水平 α 下接受 H_0.

有了这两条结论就能方便地确定是否拒绝 H_0。这种假设检验方法称为**p 值法**。

例 (8.2 节例 1)

某种元件的寿命 X(以 h 计) 服从正态分布 $N(\mu,\sigma^2)$，μ，σ^2 均未知。现在测得 16 只元件的寿命如下:

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

问是否有理由认为元件的平均寿命大于 225 h ? 用 p 值法检验假设

$$H_0 : \mu \le \mu_0 = 225, \quad H_1 : \mu > \mu_0, \quad \alpha = 0.05.$$

解: 用 t 检验法。检验统计量 $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ 的样本观察值为 $t_0 = 0.6685$。用计算机求得（即关于 $t(n-1)$ 概率密度函数的数值积分，事实上查询 t 分布的分位数表也能获得足够的信息）

$$p\text{值} = P_{\mu_0}\{T \ge t_0\} = 0.2570 > \alpha = 0.05,$$

故接受 H_0。

p 值法的优点：比临界值法给出了有关拒绝域的更多信息。

用临界值法做假设检验时，例如取 $\alpha = 0.05$ 时拒绝 H_0，再取 $\alpha = 0.01$ 时也拒绝 H_0，但我们不知道将 α 再降低一些是否也要拒绝 H_0。相反，p 值法给出了拒绝 H_0 的最小显著性水平。

延伸阅读：

- 书本 P.212–213 例 2、4；
- 书本 P.213 最后两段话对 p 值的解读：反对原假设 H_0 的依据强度。

The End

本章作业：TBA...