第 8 章-假设检验

李孟棠, 朱彬

中山大学-智能工程学院

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

2021 年 7 月 1 日

章节概述

知识点:

1. 假设检验
2. 正态总体均值的假设检验
3. 正态总体方差的假设检验

8.1 假设检验

在总体分布函数完全未知、或已知形式而不知具体参数的情况下，我们提出某些关于总体的假设。

e.g. 提出总体服从泊松分布的假设、或对正态总体提出数学期望等于 μ_0 的假设。

我们需要根据样本，接受或者拒绝这些假设。

例 (1)

某车间用一台包装机包装葡萄糖。袋装糖的净重是一个随机变量，服从正态分布。当机器正常时，均值为 $\mu_0 = 0.5$ kg，标准差 $\sigma_0 = 0.015$ kg。某日开工后为检验包装机是否正常，随机抽取包装的 9 袋糖，净重如下（kg），问机器是否正常？

$$0.497 \quad 0.506 \quad 0.518 \quad 0.524 \quad 0.498 \quad 0.511 \quad 0.520 \quad 0.515 \quad 0.512$$

解：以 μ, σ 表示这一天袋装糖的净重总体 X 的均值和标准差，由于长期实践表明标准差比较稳定，我们就假设 $\sigma = 0.015$，因此
$X \sim N(\mu, 0.015)$，这里 μ 未知。这里，我们根据样本来判断 $\mu = 0.5$ 还是 $\mu \neq 0.5$。我们提出两个对立的假设：

$$\begin{aligned} H_0: &\quad \mu = \mu_0 = 0.5 \\ H_1: &\quad \mu \neq \mu_0 \end{aligned}$$

利用已知样本作出决策，接受 H_0(拒绝 H_1)，或者接受 H_1(拒绝 H_0).

思路：

1 涉及总体均值 μ，想到样本均值 $\overline{X}$ 为 μ 的无偏估计，$\overline{X}$ 的观察值 $\overline{x}$ 的大小在一定程度上反应 μ 的大小。

2 如果 H_0 为真，观察值 $\overline{x}$ 与 μ_0 的偏差 $|\overline{x}-\mu_0|$ 一般不应该太大。如果过大，则应该拒绝 H_0。

3 当 H_0 为真时 $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$。衡量 $|\overline{x}-\mu_0|$ 的大小可以变成衡量 $\left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|$ 的大小。

4 基于上面的想法，我们可以选定一个正数 k，当观察值 $\overline{x}$ 满足 $\left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq k$ 时拒绝 H_0，反之接受 H_0。

5 然而，做出决策的依据是一个样本，当实际上 H_0 为真时仍可能做出拒绝 H_0 的决策（这种可能性无法消除，因为样本每次不同）。这种“弃真错误”，叫做第 I 类错误，概率记为

$$P\{\text{当 } H_0 \text{ 为真拒绝 } H_0\}\text{或}P_{\mu_0}\{\text{拒绝 } H_0\}\text{或}P_{\mu\in H_0}\{\text{拒绝 } H_0\}$$

记号 $P_{\mu_0}\{\sim\}$ 表示参数 μ 取 μ_0 时，事件 $\{\sim\}$ 的概率。记号 $P_{\mu\in H_0}\{\sim\}$ 表示参数 μ 取 H_0 规定的值时，事件 $\{\sim\}$ 的概率。

思路:

6 我们无法排除这类错误，自然希望将出错概率控制到较低的水平，即给定一个较小的数 $0<\alpha<1$，使得

$$P\{\text{当 } H_0 \text{ 为真拒绝 } H_0\} \le \alpha \tag{1}$$

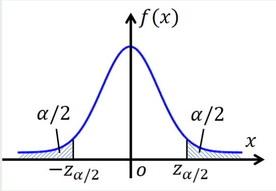
7 为了确定常数 k，考虑统计量 $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$。由于只允许犯这类错误的概率最大为 α，因此令公式 (1) 右端取等号，即

$$P\{\text{当 } H_0 \text{ 为真拒绝 } H_0\} = P_{\mu_0}\left\{\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \ge k\right\} = \alpha$$

当 H_0 为真时，$Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$，
由标准正态分布分位点的定义可得:

$$k = z_{\alpha/2}$$

8

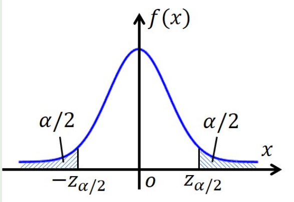


思路:

8 当 H_0 为真时，$Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

由标准正态分布分位点的定义可得:

$$k = z_{\alpha/2}$$



9 综上，若 Z 的观察值的绝对值:

$$|z| = \left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq k = z_{\alpha/2} \implies \text{拒绝}H_0$$

$$|z| = \left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right| < k = z_{\alpha/2} \implies \text{接受}H_0$$

回到本例题:

某车间用一台包装机包装葡萄糖。袋装糖的净重是一个随机变量，服从正态分布。当机器正常时，均值为 0.5 kg，标准差 0.015 kg。某日开工后为检验包装机是否正常，随机抽取包装的 9 袋糖，净重如下（kg），问机器是否正常？

$$0.497 \quad 0.506 \quad 0.518 \quad 0.524 \quad 0.498 \quad 0.511 \quad 0.520 \quad 0.515 \quad 0.512$$

解：取 $\alpha = 0.05$，则 $k = z_{\alpha/2} = 1.96$。
又已知 $n = 9, \sigma = 0.015$，由样本算得 $\overline{x} = 0.511$，既有

$$\left| \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \right| = 2.2 > 1.96 \implies \text{拒绝} H_0$$

即，我们根据样本，认为这天包装机工作不正常。

逻辑解读:
上例中采用的检验法则符合实际推断原理。

因为通常 α 取得比较小，一般为 0.01 或 0.05。因而若 H_0 为真，即 $\mu_0=\mu$ 时，$\left\{\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right|\geq z_{\alpha/2}\right\}$ 是一个小概率事件。

根据实际推断原理，如果 H_0 为真，则由一次试验得到的观察值 $\overline{x}$，满足不等式 $\left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|\geq z_{\alpha/2}$ 的可能性很小；因此，如果这样的事件确实发生了，那么我们有理由怀疑假设 H_0 的正确性，从而做出拒绝 H_0 的决定。

如果出现的观察值 $\overline{x}$，满足不等式 $\left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|<z_{\alpha/2}$，此时没有理由拒绝假设 H_0，只能接受 H_0。

名词:
数 α 称为显著性水平(significance level)。
统计量 $Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 称为检验统计量(test statistic)。

定义 (双边假设检验)

前面的检验问题通常叙述称：在显著性水平α 下，检验假设

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0, \tag{2}$$

也常说成“显著性水平 α 下，针对 H_1 检验 H_0”。H_0 称为原假设或零假设，H_1 称为备择假设，指在原假设被拒绝后可供选择的假设。

我们需要根据样本，按上述检验方法，接受假设 H_0（拒绝 H_1），或拒绝 H_0（接受 H_1）。

此种类型的检验中的备择假设 H_1，表示 μ 有可能大于 μ_0 有可能小于 μ_0，称为双边备择假设，而该假设检验称为双边假设检验。

定义 (拒绝域)

当统计量 Z 取某个区域 C 中的值时，我们拒绝原假设 H_0，这个区域 C 称为拒绝域，其边界点称为临界点，如上例中的 $z = -z_{\alpha/2}$ 和 $z = z_{\alpha/2}$。

由于检验法则是根据样本做出的，因此总有可能做出错误的决策：

1. 当 H_0 为真时拒绝 H_0，即"弃真"，称为第 I 类错误，记为

$$P\{当\ H_0\ 为真拒绝\ H_0\}或P_{\mu_0}\{拒绝\ H_0\}或P_{\mu\in H_0}\{拒绝\ H_0\}$$

2. 当 H_0 为假时接受 H_0，即"取伪"，称为第 II 类错误，记为

$$P\{当\ H_0\ 不真接受\ H_0\}或P_{\mu\in H_1}\{接受\ H_0\}$$

困境：我们当然力求降低犯两种错误的概率，但是，当样本容量固定时，减小一类错误的概率，就会增加另一类错误的概率，除非增加样本容量。

一般来说，我们只控制犯第 I 类错误的概率，使得不大于 α，一般取 0.1,0.05,0.01,0.005. 这种只控制第 I 类错误犯错概率，而不考虑犯第 II 类错误概率的检验，称为显著性检验(significance test).

不同于双边假设检验，有时候我们只关心总体均值是否增大（或减小），例如，试验新工艺以提高材料的强度。这时，所考虑的总体均值应该越大越好。如果我们能判断在新工艺下总体均值较以往正常时更大，则可以考虑采用新工艺。

定义 (单边假设检验)

若进行检验假设

$$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0 \tag{3}$$

称为右边检验。若进行检验假设

$$H_0 : \mu \geq \mu_0, \quad H_1 : \mu < \mu_0 \tag{4}$$

称为左边检验，统称为单边假设检验。

单边检验的拒绝域

例 (单边检验的拒绝域)

设总体 $X \sim N(\mu, \sigma^2)$，其中 μ 未知，σ 已知，$X_1, \ldots, X_n$ 为来自 X 的样本。给定显著性水平 α，我们来求右边检验问题的拒绝域。

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

根据上面的问题描述可以看出，H_0 中的全部 μ 都比 H_1 中的 μ 要小，当 H_1 为真时，观察值 $\overline{x}$ 往往**偏大**，因此拒绝域的形式为

$$\overline{x} \ge k \quad (k \text{ 为常数})$$

确定常数 k 的方法与前面的例 1 类似：

$$P\{\text{当 } H_0 \text{ 为真时拒绝 } H_0\} = P_{\mu \in H_0}\{\overline{X} \ge k\}$$

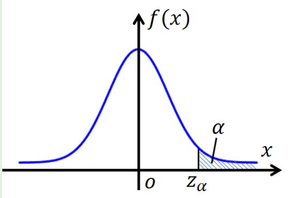
$$= P_{\mu \le \mu_0}\left\{\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \ge \frac{k - \mu_0}{\sigma/\sqrt{n}}\right\} \le P_{\mu \le \mu_0}\left\{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \ge \frac{k - \mu_0}{\sigma/\sqrt{n}}\right\}$$

要控制 $P\{$当 H_0 为真时拒绝 $H_0\} \le \alpha$，只需令

$$P_{\mu\le\mu_0}\left\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \ge \frac{k-\mu_0}{\sigma/\sqrt{n}}\right\} = \alpha \qquad (5)$$

由于 $Z := \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，可知上式 (5) 中

$$\frac{k-\mu_0}{\sigma/\sqrt{n}} = z_\alpha \implies k = \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$$



因此问题 (3) 的拒绝域为

$$\overline{x} \ge \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$$

即

$$z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} \ge z_\alpha \qquad (6)$$

总结

右边检验问题 (3):

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

的拒绝域为

$$z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \ge z_\alpha \tag{7}$$

类似的，左边检验问题 (4):

$$H_0: \mu \ge \mu_0, \quad H_1: \mu < \mu_0$$

的拒绝域为

$$z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \le -z_\alpha \tag{8}$$

其中 z_α 是标准正态分布 $N(0,1)$ 的上 α 分位点。

例 (2)

公司从生产商购买牛奶，怀疑生产商在牛奶中掺水以牟利。通过测定牛奶的冰点，可以检测是否掺水。天然牛奶的冰点温度近似服从正态分布，均值 $\mu_0 = -0.545°\mathrm{C}$，标准差 $\sigma = 0.008°\mathrm{C}$。牛奶掺水可以使冰点温度升高而接近于水的冰点（$0°\mathrm{C}$）。测得生产商提交的 5 批牛奶的冰点温度，其均值为 $\overline{x} = -0.535°\mathrm{C}$，问是否可以认为生产商的牛奶掺水？取 $\alpha = 0.05$。

解：按题意需进行检验假设

$$
\begin{aligned}
H_0 &: \mu \le \mu_0 = -0.545 \quad (\text{牛奶未掺水}) \\
H_1 &: \mu > \mu_0 = -0.545 \quad (\text{牛奶掺水})
\end{aligned}
$$

此为右边检验问题，其拒绝域如公式(7)所示，即

$$
z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \ge z_\alpha = 1.645
$$

而我们的

$$
z = \frac{-0.535 - (-0.545)}{0.008/\sqrt{5}} = 2.7951 > 1.645
$$

在拒绝域中，因此拒绝 H_0 接受 H_1，即认为生产商掺水了。

8.2 正态总体均值的假设检验

（一）单个总体 $N(\mu, \sigma^2)$ 均值 μ 的检验

（二）两个正态总体均值差的检验（t 检验）

（三）基于成对数据的检验（t 检验）

单个总体 $N(\mu,\sigma^2)$ 均值 μ 的检验

设总体 $X\sim N(\mu,\sigma^2)$，其中 μ,σ 未知。给定显著性水平 α，求检验问题

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

的拒绝域。

1. σ^2 已知，关于 μ 的检验（Z 检验）
   在上一节中，我们讨论过该检验问题：公式(2)、(3)、(4)，我们都是利用统计量 $Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 来确定拒绝域。这样的检验方法称为Z 检验法。
2. σ^2 未知，关于 μ 的检验（t 检验）
   此时 σ 未知，无法采用统计量 $Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 来确定拒绝域。
   考虑到 S^2 是 σ^2 的无偏估计，我们用 S 来代替 σ，构造新的统计量

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

   当观察值 $|t|=\left|\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right|$ 过大时就拒绝 H_0，拒绝域为

$$|t|=\left|\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right|\geq k$$

σ 未知，无法采用统计量 $Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 来确定拒绝域。
考虑到 S^2 是 σ^2 的无偏估计，我们用 S 来代替 σ，构造新的统计量

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

当观察值 $|t|=\left|\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right|$ 过大时就拒绝 H_0，拒绝域为

$$|t|=\left|\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right|\geq k$$

由第 6 章第 3 节定理 3知，当 H_0 为真时，

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

所以由

$$P\{\text{当 } H_0 \text{ 为真时拒绝 } H_0\}=P_{\mu_0}\left\{\left|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right|\geq k\right\}=\alpha$$

所以由

$$P\{\text{当 } H_0 \text{ 为真时拒绝 } H_0\} = P_{\mu_0}\left\{\left|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right| \geq k\right\} = \alpha$$

得 $k = t_{\alpha/2}(n-1)$，即拒绝域为

$$|t| = \left|\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right| \geq t_{\alpha/2}(n-1) \tag{9}$$

对于正态总体 $N(\mu, \sigma^2)$，当 σ^2 未知时，关于 μ 的单边检验的拒绝域在表8.1中给出。

上述利用 t 统计量得出的检验法称为t 检验法。

表: 8.1-1: 正态总体均值、方差的检验法，显著性水平 α

	原假设 H_0	检验统计量	备择假设 H_1	拒绝域
1	$\mu \le \mu_0$ $\mu \ge \mu_0$ $\mu = \mu_0$ (σ^2已知)	$Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \ne \mu_0$	$z \ge z_\alpha$ $z \le -z_\alpha$ $\|z\| \ge z_{\alpha/2}$
2	$\mu \le \mu_0$ $\mu \ge \mu_0$ $\mu = \mu_0$ (σ^2未知)	$T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \ne \mu_0$	$t \ge t_\alpha(n-1)$ $t \le -t_\alpha(n-1)$ $\|t\| \ge t_{\alpha/2}(n-1)$

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，问是否有理由认为元件的平均寿命大于 225 h？

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

解：按题意需进行检验

$$H_0 : \mu \leq \mu_0 = 225, \quad H_1 : \mu > \mu_0 = 225$$

取 $\alpha = 0.05$，由表 8.1 可知此检验问题的拒绝域为

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \geq t_\alpha(n-1)$$

现在 $n = 16, t_{0.05}(15) = 1.7531, \overline{x} = 241.5, s = 98.7259$，即有

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = 0.6685 < 1.7531$$

t 没有落在拒绝域内，因此接受 H_0，认为元件的平均寿命不大于 225 h.

我们还可以用 t 检验法来检验具有相同方差的两个正态总体均值差的假设。设 $X_1, \dots, X_{n_1}$ 和 $Y_1, \dots, Y_{n_2}$ 分别是来自正态总体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 的样本 $(\mu_1, \mu_2, \sigma^2$ 均未知)，且这两个样本相互独立。设 $\overline{X}, \overline{Y}$ 为对应的样本均值，S_1^2, S_2^2 为样本方差。现在来求解检验问题：

$$H_0: \mu_1 - \mu_2 = \delta, \quad H_1: \mu_1 = \mu_2 \neq \delta \quad (\delta \text{ 已知})$$

可以构造 t 统计量

$$T = \frac{(\overline{X} - \overline{Y}) - \delta}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

由第 6 章第 3 节定理 4知，当 H_0 为真时，$T \sim t(n_1 + n_2 - 2)$。与单个总体的 t 检验法相似，其拒绝域的形式为

$$\left| \frac{(\overline{x} - \overline{y}) - \delta}{s_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \geq k$$

由

$$P\{当\ H_0\ 为真时拒绝\ H_0\} = P_{\mu_1-\mu_2=\delta}\left\{\left|\frac{(\overline{X}-\overline{Y})-\delta}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| \geq k\right\} = \alpha$$

可得 $k = t_{\alpha/2}(n_1+n_2-2)$。于是拒绝域为

$$|t| = \frac{|(\overline{x}-\overline{y})-\delta|}{s_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \geq t_{\alpha/2}(n_1+n_2-2) \tag{10}$$

关于均值差的两个单边检验问题的拒绝域在表 8.1 中给出，常用 $\delta = 0$。

当两个正态总体的方差 σ_1, σ_2 均已知 (不一定相等) 时，我们可以用*Z* 检验法来验证两个正态总体均值差的假设问题。

表: 8.1-2: 正态总体均值、方差的检验法，显著性水平 α

	原假设 H_0	检验统计量	备择假设 H_1	拒绝域
3	$\mu_1-\mu_2\le\delta$ $\mu_1-\mu_2\ge\delta$ $\mu_1-\mu_2=\delta$ $(\sigma_1^2,\sigma_2^2$已知$)$	$Z=\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$\mu_1-\mu_2>\delta$ $\mu_1-\mu_2<\delta$ $\mu_1-\mu_2\ne\delta$	$z\ge z_\alpha$ $z\le -z_\alpha$ $\lvert z\rvert\ge z_{\alpha/2}$
4	$\mu_1-\mu_2\le\delta$ $\mu_1-\mu_2\ge\delta$ $\mu_1-\mu_2=\delta$ $(\sigma_1^2=\sigma_2^2=\sigma^2,$ σ^2未知$)$	$T=\frac{\overline{X}-\overline{Y}-\delta}{S_w\sqrt{\frac{1}{n_1}+\frac{2}{n_2}}}$ $S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$	$\mu_1-\mu_2>\delta$ $\mu_1-\mu_2<\delta$ $\mu_1-\mu_2\ne\delta$	$t\ge t_\alpha(n_1+n_2-2)$ $t\le -t_\alpha(n_1+n_2-2)$ $\lvert t\rvert\ge t_{\alpha/2}(n_1+n_2-2)$

例 (2)

用两种方法 A 和 B 测定冰自 $-0.72°$C 转变为 $0°$C 的水的融化热 (以 cal/g 计)，测得以下数据：

方法 A	79.98	80.04	80.02	80.04	80.03	80.03		
	80.04	79.97	80.05	80.03	80.02	80.00	80.02	
方法 B	80.02	79.94	79.98	79.97	79.97	80.03	79.95	79.97

设这两个样本相互独立且分别来自正态总体 $N(\mu_1, \sigma^2), N(\mu_2, \sigma^2), \mu_1, \mu_2, \sigma^2$ 均未知。试在显著性水平为 0.05 下进行检验假设：

$$H_0 : \mu_1 - \mu_2 \leq 0, \quad H_1 : \mu_1 - \mu_2 > 0$$

解: 因为是比较两组正态分布数据，且 $\sigma_1 = \sigma_2 = \sigma$ 未知，μ_1, μ_2 未知，因此采用表 8.1第 4 栏的方法：

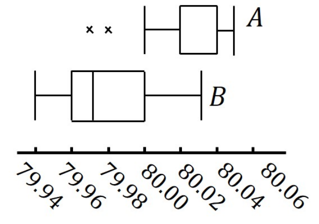
$$n_1 = 13, \ \overline{x}_A = 80.02, \ s_A^2 = 0.024^2$$
$$n_2 = 8, \ \overline{x}_B = 79.98, \ s_B^2 = 0.031^2$$
$$s_w^2 = \frac{12 \times s_A^2 + 7 \times s_B^2}{19} = 0.0007178$$
$$t = \frac{\overline{x}_A - \overline{x}_B}{s_w\sqrt{1/13 + 1/8}} = 3.323 > t_{0.05}(13 + 8 - 2) = 1.7291$$

例 (2 续)

因为是比较两组正态分布数据，且 $\sigma_1 = \sigma_2 = \sigma$ 未知，μ_1, μ_2 未知，因此采用表 8.1的 4 方法：

$$
\begin{aligned}
n_1 &= 13,\ \overline{x}_A = 80.02,\ s_A^2 = 0.024^2 \\
n_2 &= 8,\ \overline{x}_B = 79.98,\ s_B^2 = 0.031^2 \\
s_w^2 &= \frac{12\times s_A^2 + 7\times s_B^2}{19} = 0.0007178 \\
t &= \frac{\overline{x}_A - \overline{x}_B}{s_w\sqrt{1/13+1/8}} = 3.323 > t_{0.05}(13+8-2) = 1.7291
\end{aligned}
$$

故拒绝 H_0，认为 A 方法比 B 方法测得的融化热要大。



基于成对数据的检验（t 检验）

有时为了比较两种产品、两种仪器、两种方法等的差异，常常在相同的条件下做对比实验，得到一批成对的观察值。然后分析观察数据做出推断。这种方法称为逐对比较法。

例 (3)

有两台光谱仪 I_x, I_y 用来测量材料中某种金属的含量，为鉴定它们的测量结果有无显著差异，制备了 9 件试块（它们的成分、金属含量、均匀性等均各不相同），现在分别用这两台仪器对每一试块测量一次，得到 9 对观察值如下，问能否认为这两台仪器的测量结果有显著差异（$\alpha = 0.01$）？

$x(\%)$	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
$y(\%)$	0.10	0.21	0.52	0.32	0.78	0.59	0.68	0.77	0.89
$d = x - y(\%)$	0.10	0.09	-0.12	0.18	-0.18	0.11	0.12	0.13	0.11

问：可以使用表 8.1 中的第 4 栏来进行检验假设吗？

思路:

1. 本题中的数据是成对的，即对同一试块测出的一对数据。
2. 我们看到一对与另一对之间的差异是由各种因素，如材料成分、金属含量、均匀性等因素引起的，因此不能将仪器 I_x 对 9 个试块的测量结果看成是同分布随机变量的观察值。
3. 因此表中第一行不能看成是一个样本的样本值。同理，表中第二行也不能看成是一个样本的样本值。
4. 再者，对每一对数据而言，它们是同一试块用不同仪器测得的结果。因此它们不是两个独立的随机变量的观察值。
5. 同一对中两个数据的差异则可以看成是仅由这两台仪器性能差异所引起的，这样，局限于各对中两个数据来比较就能排出种种其他因素，而只考虑单独由仪器的性能所产生的的影响。

方法:

1 一般，设有 n 对相互独立的观察结果：$(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$。

2 令 $D_1 = X_1 - Y_1, \cdots, D_n = X_n - Y_n$，则 $D_1, D_2, \cdots, D_n$ 相互独立。

3 又由于 $D_1, D_2, \cdots, D_n$ 由同一因素引起，因此可认为它们服从同一分布。

4 假设 $D_i \sim N(\mu_D, \sigma_D^2)$，$D_1, D_2, \cdots, D_n$ 构成正态总体 $N(\mu_D, \sigma_D^2)$ 的一个样本，其中 μ_D, σ_D^2 未知。我们需要基于这一样本检验假设：

$$\begin{array}{lll} (1) & H_0: \mu_D = 0, & H_1: \mu_D \neq 0 \\ (2) & H_0: \mu_D \leq 0, & H_1: \mu_D > 0 \\ (3) & H_0: \mu_D \geq 0, & H_1: \mu_D < 0 \end{array}$$

5 分别计 $D_1, D_2, \cdots, D_n$ 的样本均值和样本方差的观察值为 $\bar{d}, s_D^2$，按表 8.1的第 2 栏关于单个正态总体均值的 t 检验进行检验，则关于 (1)(2)(3) 的拒绝域分别为

$$|t| = \left|\frac{\bar{d}}{s_d/\sqrt{n}}\right| \geq t_{\alpha/2}(n-1)$$

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \geq t_\alpha(n-1)$$

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \leq -t_\alpha(n-1)$$

回到该例题

$x(\%)$	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
$y(\%)$	0.10	0.21	0.52	0.32	0.78	0.59	0.68	0.77	0.89
$d = x - y(\%)$	0.10	0.09	-0.12	0.18	-0.18	0.11	0.12	0.13	0.11

解：我们要检验两台仪器是否有差异，即进行如下假设：

$$(1) \quad H_0 : \mu_D = 0, \quad H_1 : \mu_D \neq 0$$

现在 $n = 9, t_{\alpha/2}(8) = t_{0.005}(8) = 3.3554$，可算得

$$|t| = \left| \frac{\overline{d}}{s_d/\sqrt{n}} \right| = 1.467 < 3.3554$$

没落在拒绝域内，因此接受 H_0 假设，即认为两台仪器没有差异。

例 (4)

做以下实验以比较人对红光或绿光的反应时间 (以 s 计)。实验在点亮红光或绿光的同时启动计时器，要求受试者见到红光或绿光时按下按钮以切断计时器，从而测得反应时间。结果如下

红光 x	0.30	0.23	0.41	0.53	0.24	0.36	0.38	0.51
绿光 y	0.43	0.32	0.58	0.46	0.27	0.41	0.38	0.61
$d = x - y$	-0.13	-0.09	-0.17	0.07	-0.03	-0.05	0.00	-0.10

设 $D_i = X_i - Y_i$ 是来自正态总体 $N(\mu_D, \sigma_D^2)$ 的样本，μ_D, σ_D^2 未知。现在检验左边假设 (显著性水平取 $\alpha = 0.05$)

$$H_0 : \mu_D \geq 0, \quad H_1 : \mu_D < 0.$$

解：现有 $n = 8, \overline{d} = -0.0625, s_D = 0.0765$，可知

$$\frac{\overline{d}}{s_d/\sqrt{n}} = -2.311 < -t_{0.05}(7) = -1.8946$$

落在拒绝域内，此时拒绝 H_0，即认为人对红光反应时间小于对绿光反应时间。

8.3 正态总体方差的假设检验

（一）单个总体的情况

（二）两个正态总体的情况

单个总体的情况

设总体 $X \sim N(\mu, \sigma^2)$，μ, σ^2 均未知，$X_1, X_2, \cdots, X_n$ 是来自 X 的样本。进行双边检验假设（显著性水平 α）

$$H_0 : \sigma^2 = \sigma_0^2, \quad H_1 : \sigma^2 \neq \sigma_0^2 \quad (\sigma_0^2\text{已知}).$$

由于 S^2 是 σ^2 的无偏估计，当 H_0 为真时，观察值 s^2 与 σ_0^2 的比值一般来说在 1 附近摆动，不应过分大于 1 或小于 1. 由第 6 章第 3 节定理 3可知，当 H_0 为真时

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

我们取统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

上述检验问题的拒绝域为

$$\frac{(n-1)S^2}{\sigma_0^2} \leq k_1 \quad \text{或} \quad \frac{(n-1)S^2}{\sigma_0^2} \geq k_2$$

k_1, k_2 的值由下式确定

$$P_{H_0}\{拒绝\ H_0\} = P_{\sigma_0^2}\left\{\left\{\frac{(n-1)S^2}{\sigma_0^2} \le k_1\right\} \bigcup \left\{\frac{(n-1)S^2}{\sigma_0^2} \ge k_2\right\}\right\} = \alpha$$

为便于计算，上式常常化为

$$P_{\sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma_0^2} \le k_1\right\} = \frac{\alpha}{2}, \quad P_{\sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma_0^2} \ge k_2\right\} = \frac{\alpha}{2}$$

故得 $k_1 = \chi^2_{1-\alpha/2}(n-1), k_2 = \chi^2_{\alpha/2}(n-1)$，于是得到拒绝域

$$\frac{(n-1)S^2}{\sigma_0^2} \le \chi^2_{1-\alpha/2}(n-1) \quad 或 \quad \frac{(n-1)S^2}{\sigma_0^2} \ge \chi^2_{\alpha/2}(n-1) \tag{11}$$

这种方法叫做χ^2 检验法，对应表 8.1 的第 5 栏。

表: 8.1-3: 正态总体均值、方差的检验法，显著性水平 α

	原假设 H_0	检验统计量	备择假设 H_1	拒绝域
5	$\sigma^2 \le \sigma_0^2$ $\sigma^2 \ge \sigma_0^2$ $\sigma^2 = \sigma_0^2$ (μ未知)	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \ne \sigma_0^2$	$\chi^2 \ge \chi_\alpha^2(n-1)$ $\chi^2 \le \chi_{1-\alpha}^2(n-1)$ $\chi^2 \ge \chi_{\alpha/2}^2(n-1)$或 $\chi^2 \le \chi_{1-\alpha/2}^2(n-1)$
6	$\sigma_1^2 \le \sigma_2^2$ $\sigma_1^2 \ge \sigma_2^2$ $\sigma_1^2 = \sigma_2^2$ (μ_1, μ_2未知)	$F = \frac{S_1^2}{S_2^2}$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \ne \sigma_0^2$	$F \ge F_\alpha^2(n_1-1, n_2-1)$ $F \le F_{1-\alpha}^2(n_1-1, n_2-1)$ $F \ge F_{\alpha/2}^2(n_1-1, n_2-1)$或 $F \le F_{1-\alpha/2}^2(n_1-1, n_2-1)$
7	$\mu_D \le 0$ $\mu_D \ge 0$ $\mu_D = 0$ 成对数据	$T = \frac{\overline{D}-0}{S_D/\sqrt{n}}$	$\mu_D > 0$ $\mu_D < 0$ $\mu_D \ne 0$	$t \ge t_\alpha(n-1)$ $t \le -t_\alpha(n-1)$ $\|t\| \ge t_{\alpha/2}(n-1)$

设总体 $X \sim N(\mu, \sigma^2)$，μ, σ^2 均未知，$X_1, X_2, \cdots, X_n$ 是来自 X 的样本，进行右边检验假设（显著性水平 α）

$$H_0 : \sigma^2 \leq \sigma_0^2, \quad H_1 : \sigma^2 > \sigma_0^2, \quad (\sigma_0^2\text{已知}) \tag{12}$$

因为 H_0 中的全部 σ^2 都要比 H_1 中的 σ^2 要小，当 H_1 为真时，S^2 的观察值 s^2 往往偏大，因此拒绝域形式为

$$s^2 \geq k$$

k 的值按照以下方式确定

$$\begin{aligned} P\{\text{当 } H_0 \text{ 为真时拒绝 } H_0\} = P_{\sigma^2 \leq \sigma_0^2}\{S^2 \geq k\} &= P_{\sigma^2 \leq \sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)k}{\sigma_0^2}\right\} \\ &\leq P_{\sigma^2 \leq \sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma^2} \geq \frac{(n-1)k}{\sigma_0^2}\right\} \end{aligned}$$

其中最后一个不等式来自于零假设 $\sigma^2 \leq \sigma_0^2$。

因此，为了控制 $P\{\text{当 } H_0 \text{ 为真时拒绝 } H_0\} \leq \alpha$，只需令

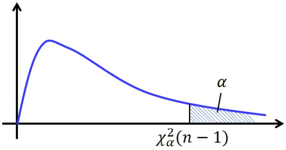
$$P_{\sigma^2 \leq \sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma^2} \geq \frac{(n-1)k}{\sigma_0^2}\right\} = \alpha \tag{13}$$

因为

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

等式 (13) 的解为

$$\frac{(n-1)k}{\sigma_0^2} = \chi_\alpha^2(n-1) \quad \text{(如右图所示)}$$



于是 $k = \frac{\sigma_0^2}{n-1}\chi_\alpha^2(n-1)$，得到右边检验问题 (12) 的拒绝域为

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \geq \chi_\alpha^2(n-1) \tag{14}$$

类似地，左边检验问题：

$$H_0 : \sigma^2 \geq \sigma_0^2, \quad H_1 : \sigma^2 < \sigma_0^2 \quad (\sigma_0^2\text{已知})$$

的拒绝域为

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \leq \chi_{1-\alpha}^2(n-1) \tag{15}$$

例 (1)

某工厂生产的某型电池，其寿命 (以 h 计) 长期以来服从方差 $\sigma^2 = 5000$ 的正态分布。现有一批这种电池，从它的生产情况来看，寿命的波动性（即方差）有所改变。现随机取 26 只电池，测出其寿命的样本方差 $s^2 = 9200$。问根据这一数据能否推断这批电池寿命的波动性较以往有显著的变化（取 $\alpha = 0.02$)。

解：本题需要进行检验

$$H_0 : \sigma^2 = 5000, \quad H_1 : \sigma^2 \neq 5000$$

现在 $n = 26,\ \sigma_0^2 = 5000$

$$\chi^2_{\alpha/2}(n-1) = \chi^2_{0.01}(25) = 44.314, \quad \chi^2_{1-\alpha/2}(n-1) = \chi^2_{0.99}(25) = 11.524$$

由公式 (11) 可知拒绝域为

$$\frac{(n-1)s^2}{\sigma_0^2} \geq 44.314 \quad \text{或} \quad \frac{(n-1)s^2}{\sigma_0^2} \leq 11.524$$

带入观察值 $s^2 = 9200$ 得

$$\frac{(n-1)s^2}{\sigma_0^2} = 46 \geq 44.314$$

所以拒绝 H_0，认为这批电池寿命的波动较以往确实有显著变化。

两个总体的情况

设 $X_1, X_2, \cdots, X_{n_1}$ 是来自总体 $N(\mu_1, \sigma_1^2)$ 的样本，$Y_1, Y_2, \cdots, Y_{n_2}$ 是来自总体 $N(\mu_2, \sigma_2^2)$ 的样本，两样本相互独立，其样本方差分别为 S_1^2, S_2^2，且设 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 均未知。现在检验如下假设（显著性水平 α）：

$$H_0 : \sigma_1^2 \leq \sigma_2^2, \quad H_1 : \sigma_1^2 > \sigma_2^2 \tag{16}$$

当 H_0 为真时，

$$E(S_1^2) = \sigma_1^2 \leq \sigma_2^2 = E(S_2^2)$$

当 H_1 为真时，

$$E(S_1^2) = \sigma_1^2 > \sigma_2^2 = E(S_2^2)$$

此时观察值 S_1^2/S_2^2 有偏大的趋势，故拒绝域为

$$\frac{s_1^2}{s_2^2} \geq k$$

k 的值按照以下方式确定（最后的 $\leq$ 是因为在零假设下 $\sigma_1^2/\sigma_2^2 \leq 1$）

$$P\{\text{当 } H_0 \text{ 为真时拒绝 } H_0\} = P_{\sigma_1^2 \leq \sigma_2^2}\left\{\frac{S_1^2}{S_2^2} \geq k\right\} \leq P_{\sigma_1^2 \leq \sigma_2^2}\left\{\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \geq k\right\}$$

因此，要控制 $P\{当\ H_0\ 为真时拒绝\ H_0\} \le \alpha$，只需令

$$P_{\sigma_1^2 \le \sigma_2^2}\left\{\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \ge k\right\} = \alpha \tag{17}$$

由第 6 章第 3 节定理 4可知

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

求解等式 (17) 可得 $k = F_\alpha(n_1-1, n_2-1)$，即检验问题 (16) 的拒绝域为

$$F = \frac{s_1^2}{s_2^2} \ge F_\alpha(n_1-1, n_2-1) \tag{18}$$

上述检验方法叫做*F* 检验法，对应表 8.1 的第 6 栏。

例 (2(第 2 节例 2))

方法 A 和 B 测定冰自 $-0.72°$C 转变为 $0°$C 的水的融化热 (以 cal/g 计)，测得数据：

方法 A	79.98	80.04	80.02	80.04	80.03	80.03		
	80.04	79.97	80.05	80.03	80.02	80.00	80.02	
方法 B	80.02	79.94	79.98	79.97	79.97	80.03	79.95	79.97

设这两个样本相互独立且分别来自正态总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$，其中 $\mu_1, \mu_2, \sigma_1, \sigma_2$ 均未知。注意：我们在课件 25 页（以及教材中相应部分）假设两个总体具有相同方差。这里我们对此假设进行检验（$\alpha = 0.01$）：

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

解：现在有 $n_1 = 13, n_2 = 8, \alpha = 0.01$，公式 (18) 得到拒绝域为

$$s_1^2/s_2^2 \geq F_{\alpha/2}(n_1 - 1, n_2 - 1) = F_{0.005(12,7)} = 8.18$$
$$s_1^2/s_2^2 \leq F_{1-\alpha/2}(n_1 - 1, n_2 - 1) = 1/F_{0.005(7,12)} = 0.18$$

现在有样本方差观察值 $s_1^2 = 0.024^2, s_2^2 = 0.031^2, s_1^2/s_2^2 = 0.60$，不在拒绝域内，因此接受 H_0，即认为两总体方差相等，从而验证了之前的假设。

章节结束