第八章　假设检验

统计推断的另一种重要形式是假设检验.概括起来讲,所谓假设检验就是根据样本中的信息来检验总体的分布参数或分布形式具有指定的特征.例如,对于一个正态总体,我们通过样本来推断该总体的均值是否等于给定值 μ_0,就是一个最简单最重要的假设检验问题.

本章分为五节,在 §8.1 中将首先引进假设检验中的一些重要的基本概念.在实用中,正态总体是最重要的研究对象.于是,在 §8.2 和 §8.3 中,我们将详细讨论正态总体均值和方差的检验.在实际应用中,我们往往并不是一开始就能确定总体的分布形式.比如,在连续型分布的情形,时常对总体是正态分布,还是指数分布或者是其他分布并无完全把握,这就要通过样本来检验总体的具体分布形式.§8.4 要讨论的拟合优度检验就是用来解决这个问题的.最后一节把拟合优度检验应用于列联表.列联表在社会、医学调查中具有广泛的应用.

§8.1　基本概念

我们通过一个例子,来引进假设检验中的一些重要概念.

例 8.1.1　某工厂生产 10 欧姆的电阻.根据以往生产的电阻实际情况,可以认为其电阻值服从正态分布,标准差 $\sigma = 0.1$.现在随机抽取 10 个电阻,测得它们的电阻值为

9.9,　10.1,　10.2,　9.7,　9.9,　9.9,　10,　10.5,　10.1,　10.2.

问,从这些样本我们能否认为该厂生产的电阻的平均值为 10 欧姆?

记 X 为该厂生产的电阻的测量值.根据假设,$X \sim N(\mu, \sigma^2)$,这里 $\sigma = 0.1$.我们想通过样本推断的是总体均值 μ 是不是等于 10 欧姆.这个问题,在统计学上可以做如下表述.

我们有一个假设

$$H_0: \mu = 10.$$

现在要通过样本去检验这个假设是否成立.这个假设的对立面是 $H_1: \mu \neq 10$.把它们合写在一起,就是

$$H_0: \mu = 10 \longleftrightarrow H_1: \mu \neq 10. \qquad (8.1.1)$$

在数理统计中,我们把“$H_0:\mu=10$”称为“原假设”或“零假设”(这里“零假设”是从英文“null hypothesis”一词而来),而把“$H_1:\mu\neq 10$”称为“对立假设”或“备择假设”.

我们知道,样本均值 $\bar{X}$ 是总体 μ 的一个良好估计.因此,如果 $\mu=10$,也就是说原假设成立,那么 $|\bar{X}-10|$ 应该比较小.~~反过来~~,若原假设不成立,则它就应该比较大.因此,$|\bar{X}-10|$ 的大小可以用来检验原假设是否成立.直观上合理的检验是当 $|\bar{X}-10|<c$ 时,我们就接受原假设 H_0,而当 $|\bar{X}-10|\geqslant c$ 时,我们就拒绝原假设 H_0.这里的问题是,我们如何确定常数 c 呢?

根据基本定理(见定理 6.4.1):$\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$,即

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

对现在的情形:$n=10,\sigma^2=0.1^2$,于是,当原假设成立时

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{\bar{X}-10}{0.1/\sqrt{10}}\sim N(0,1),\tag{8.1.2}$$

对给定的 $\alpha,0<\alpha<1$,根据分位点的定义有

$$P\left\{\left|\frac{\bar{X}-10}{0.1/\sqrt{10}}\right|\geqslant Z_{\alpha/2}\right\}=\alpha,\tag{8.1.3}$$

即

$$P\{|\bar{X}-10|\geqslant(0.1/\sqrt{10})Z_{\alpha/2}\}=\alpha.$$

这样,我们就得到了 c 的值:$c=\frac{0.1}{\sqrt{10}}Z_{\alpha/2}$.这样确定 c 值的理由,涉及到检验的显著性水平,我们将在后面的讨论中作进一步说明.

由此我们得到如下检验.

(1) 若

$$|\bar{X}-10|\geqslant\frac{0.1}{\sqrt{10}}Z_{\alpha/2}=0.032Z_{\alpha/2},\tag{8.1.4}$$

则认为样本均值 $\bar{X}$ 与总体均值的假定值 10 相距太大,所以应该拒绝原假设.

(2) 若

$$|\bar{X}-10|<0.032Z_{\alpha/2},\tag{8.1.5}$$

我们就认为样本均值与总体均值的假定值 10 比较接近,于是,应该接受原假

设.

我们把(8.1.2)中的$(\bar{X}-10)\Big/(0.1/\sqrt{10})$称为检验统计量，而把(8.1.4)所定义的区域

$$|\bar{X}-10|\geqslant 0.032Z_{\alpha/2},$$

称为该检验的拒绝域.

当我们检验一个假设H_0时,有可能犯以下两类错误之一.第一,H_0是正确的,但被我们拒绝了,这就犯了“弃真”的错误,即抛弃了正确假设.第二,H_0是不正确的,但被我们接受了,这就犯了“采伪”的错误,即采用了伪假设.因为检验统计量总是随机的,所以,我们总是以一定的概率犯以上两类错误.在统计学中,把犯第一类错误的概率称为假设检验的显著性水平,简称水平.应用上,一般是采用事先把显著性水平固定下来的方法.至于犯第二类错误的概率的计算超出了本书的范围,因此不作讨论.

引进了显著性水平的概念之后,我们回过头来解释确定c值的方法的理由.假定原假设H_0正确,由(8.1.3)知,这时事件$\{|\bar{X}-10|\geqslant 0.032Z_{\alpha/2}\}$发生的概率为$\alpha$,也就是说,我们是以概率$\alpha$拒绝原假设的,因此,这时犯第一类错误的概率是$\alpha$.换句话说,该检验的显著性水平为$\alpha$.于是我们所用的确定$c$值的方法,保证了所构造的检验具有给定的显著性水平$\alpha$.

再回到前面的例子,原假设$H_0:\mu=10$的拒绝域为$|\bar{X}-10|\geqslant 0.032Z_{\alpha/2}$.由(8.1.3)知,当原假设成立时,样本落在拒绝域的概率为α.于是由(8.1.4)和(8.1.5)给出的检验犯第一类错误的概率为α,即该检验的显著性水平为α.

一般我们把显著性水平限定在一个比较小的值,通常$\alpha=0.05$或0.01.在上面的例子中,若$\alpha=0.05$,则$Z_{\alpha/2}=Z_{0.025}=1.96$,所以,原假设$H_0:\mu=10$的拒绝域为

$$|\bar{X}-10|\geqslant 0.063.$$

经计算得$\bar{X}=10.05$,$|\bar{X}-10|=0.05<0.063$,所以,我们接受原假设$\mu=10$,即认为该厂生产的电阻值的均值为10欧姆.

§8.2　正态总体均值的检验

在实际应用中,许多量都可以近似地用正态总体去刻画,因此,关于正态总体均值的检验会经常遇到.下面我们就此问题,分几种情况来讨论.

一、单个正态总体 $N(\mu,\sigma^2)$ 均值 μ 的检验

我们讨论几种常用的原假设和对立假设的情况.

1. $H_0:\mu=\mu_0 \longleftrightarrow H_1:\mu\neq\mu_0$.

假设 σ^2 已知,根据上节的讨论,我们取检验统计量

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

当 H_0 成立. $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

对给定的显著性水平 α,拒绝域可取为

$$\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| \geqslant Z_{\alpha/2},$$

即

$$|\bar{X}-\mu_0| \geqslant \frac{\sigma}{\sqrt{n}}Z_{\alpha/2}. \tag{8.2.1}$$

在应用上,σ^2 未知是常见的,此时和前面不同的是,我们需要用样本方差 $S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2\Big/(n-1)$ 来代替 σ^2,据基本定理(见定理 6.4.1)有

$$\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}.$$

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

对给定的显著性水平 α,拒绝域为

$$\left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right| \geqslant t_{n-1}\left(\frac{\alpha}{2}\right),$$

即

$$|\bar{X}-\mu_0| \geqslant \frac{S}{\sqrt{n}}t_{n-1}\left(\frac{\alpha}{2}\right). \tag{8.2.2}$$

比较(8.2.1)和(8.2.2),我们不难看出,从 σ^2 已知到未知,只需要将 $\sigma Z_{\alpha/2}$ 用 $St_{n-1}\left(\frac{\alpha}{2}\right)$来代替,所得到的检验为:若

$$|\bar{X}-\mu_0| \geqslant \frac{S}{\sqrt{n}}t_{n-1}\left(\frac{\alpha}{2}\right),$$

则拒绝原假设 $H_0:\mu=\mu_0$. 否则我们接受 H_0. 文献中常称这个检验法为 t 检验法.

原假设 $H_0:\mu=\mu_0$ 的拒绝域(8.2.1)和(8.2.2)在直观上是很合理的. 因

为样本均值 $\bar{X}$ 是总体均值 μ 的估计. 当 H_0 成立时, $|\bar{X}-\mu_0|$ 应该比较小. 如果它比较大时,我们就有理由拒绝原假设. (8.2.1) 中的 $\frac{\sigma}{\sqrt{n}}Z_{\alpha/2}$ 和(8.2.2) 中 $\frac{S}{\sqrt{n}}t_{n-1}\left(\frac{\alpha}{2}\right)$ 只是保证所得的检验具有显著性水平 α.

例 8.2.1(续例 8.1.1)　在例 8.1.1 中,假设 σ^2 未知,计算得 $S^2=0.05$. 给定显著性水平 $\alpha=0.05$,则 $t_{n-1}\left(\frac{\alpha}{2}\right)=t_9(0.025)=2.262$. 于是,拒绝域为

$$|\bar{X}-\mu_0|\geqslant\frac{S}{\sqrt{n}}t_{n-1}\left(\frac{\alpha}{2}\right)=\frac{0.224}{\sqrt{10}}\times 2.262=\frac{0.224}{3.16}\times 2.262=0.160,$$

但 $|\bar{X}-\mu_0|=|10.05-10|=0.05<0.16$,所以,在方差 σ^2 未知的情形,我们仍然是接受原假设 $\mu=10$.

2. $H_0:\mu=\mu_0 \longleftrightarrow H_1:\mu>\mu_0$.

这种形式的假设检验问题,也很有实用意义. 例如,工厂生产的产品的某项指标平均值为 μ_0. 采用了新技术或新配方后,被认为产品质量提高了,该指标的平均值 μ 应该随之上升. 我们的问题就是检验 $\mu=\mu_0$,即新技术或新配方对提高产品质量无效果,还是 $H_1:\mu>\mu_0$,即新技术或新配方确实有效,提高了产品质量.

假设 $X_1,X_2,\cdots,X_n$ 来自正态总体 $N(\mu,\sigma^2)$,方差 σ^2 未知. 由于 $\bar{X}$ 是 μ 的估计,因此,当 $\bar{X}-\mu_0$ 较大时,我们有理由认为原假设 $H_0:\mu=\mu_0$ 不成立,而对立假设 $H_1:\mu>\mu_0$ 成立. 对给定的显著性水平 α,我们知道当原假设成立时, $\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t_{n-1}$,于是拒绝域应为

$$\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\geqslant t_{n-1}(\alpha), \tag{8.2.3}$$

即

$-\frac{S}{\sqrt{n}}t_{n-1}(\alpha)$,

$$\bar{X}-\mu_0\geqslant\frac{S}{\sqrt{n}}t_{n-1}(\alpha). \tag{8.2.4}$$

所以,我们所求的检验是,当(8.2.4) 成立时,拒绝原假设,否则就接受原假设. 因为,当原假设成立时,

$$P\left\{\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\geqslant t_{n-1}(\alpha)\right\}=\alpha,$$

于是我们的检验的显著性水平为 α.

例 8.2.2　某厂生产一种工业用绳,其质量指标是绳子所承受的最大拉

力.假定该指标服从正态分布.原来该厂生产的这种绳子平均最大拉力 $\mu_0 = 15$ 公斤.现在采用了一种新的原材料,厂方称这种原材料提高了绳子的质量,也就是说绳子所承受的最大拉力比 15 公斤大了.为了检验该厂的结论是否真实,从其新产品中随机抽取 50 件,测得它们承受的最大拉力的平均值为 15.8 公斤,样本标准差 $S = 0.5$ 公斤.取显著性水平 $\alpha = 0.01$,问从这些样本看,我们能否接受厂方的结论,即新原材料是否确实提高了绳子的质量?

解　问题归结为检验如下假设

$$H_0:\mu = 15 \longleftrightarrow H_1:\mu > 15.$$

此处 $n = 50, \alpha = 0.01$,标准差 $S = 0.5$.注意到,在 t 分布表中,往往没有 $n - 1 = 49$ 对应的分位点,但由 t 分位点关于自由度 n 的单调性:$t_n(\alpha) < t_{n-1}(\alpha)$ 知 $t_{49}(0.01) < t_{45}(0.01) = 2.41$,所以,

$$\frac{S}{\sqrt{n}}t_{n-1}(\alpha) = \frac{0.5}{\sqrt{50}}t_{49}(0.01) < \frac{0.5}{\sqrt{50}}t_{45}(0.01) = 0.173.$$

但 $\bar{X} - \mu_0 = 15.8 - 15 = 0.8 > 0.173$,于是我们拒绝原假设,认为新的原材料确实提高了绳子所能承受的最大拉力.

二、两个正态总体 $N(\mu_1,\sigma_1^2)$ 和 $N(\mu_2,\sigma_2^2)$ 均值的比较

在应用上,我们经常会遇到两个正态总体均值的比较问题.譬如,欲比较甲、乙两厂生产的某种产品的质量.我们把两厂生产的产品的质量指标分别看成两个正态总体,比较它们的产品质量指标的问题,就变为比较这两个正态总体均值的问题.又如,欲考察一项新技术对提高产品质量是否有效,则把新技术实施前后生产的产品质量指标分别看成一个正态总体,这时,我们所考察的问题,就归结为检验这两个正态总体的均值是否相等的问题.

设 $X_1,X_2,\cdots,X_m$ 和 $Y_1,Y_2,\cdots,Y_n$ 分别为来自正态总体 $N(\mu_1,\sigma_1^2)$ 和 $N(\mu_2,\sigma_2^2)$ 的样本.我们要检验假设

$$H_0:\mu_1 = \mu_2 \longleftrightarrow H_1:\mu_1 \neq \mu_2.$$

首先假设 σ_1^2 和 σ_2^2 皆已知.以 $\bar{X}$ 和 $\bar{Y}$ 分别记它们的样本均值,根据定理 7.5.1 知

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} \sim N(0,1).$$

现在取上式左端为检验统计量.当 H_0 成立时

$$\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \sim N(0,1). \tag{8.2.5}$$

$\mu_1 - \mu_2 = 0$

于是,显著性水平为 α 的检验的拒绝域为

$$\frac{|\overline{X}-\overline{Y}|}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \geqslant Z_{\alpha/2},$$

即

$$|\overline{X}-\overline{Y}| \geqslant Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}. \tag{8.2.6}$$

当(8.2.6)式成立时,我们拒绝原假设,否则就接受原假设.拒绝域(8.2.6)式的直观意义是很明显的.因为 $\overline{X}$ 和 $\overline{Y}$ 分别是 μ_1 和 μ_2 的估计,因此 $\overline{X}-\overline{Y}$ 作为 $\mu_1-\mu_2$ 的估计,自然当 $|\overline{X}-\overline{Y}|$ 比较大时,原假设 $\mu_1=\mu_2$ 就不像是正确的,因此就应该拒绝它.

当 σ_1^2 和 σ_2^2 都未知时,我们假设 $\sigma_1^2=\sigma_2^2=\sigma^2$,即 σ^2 是两个正态总体的公共的,但未知的方差.记

$$S_1^2=\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\overline{X})^2,\quad S_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$$

分别为两个正态总体的样本方差.用

$$S^2=\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2} \tag{8.2.7}$$

作为 σ^2 的估计.根据定理 7.5.1 知

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t_{m+n-2}. \tag{8.2.8}$$

因此,当 H_0 成立时,

$$\frac{\overline{X}-\overline{Y}}{S\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t_{m+n-2},$$

即

$$\sqrt{\frac{mn}{m+n}}\,\frac{\overline{X}-\overline{Y}}{S} \sim t_{m+n-2}. \tag{8.2.9}$$

取左端为检验统计量,则显著性水平为 α 的检验的拒绝域为

$$\sqrt{\frac{mn}{m+n}}\frac{|\bar{X}-\bar{Y}|}{S}\geqslant t_{m+n-2}\left(\frac{\alpha}{2}\right),$$

即

$$|\bar{X}-\bar{Y}|\geqslant\sqrt{\frac{m+n}{mn}}St_{m+n-2}\left(\frac{\alpha}{2}\right). \tag{8.2.10}$$

当(8.2.10)成立时,拒绝 H_0,否则就接受 H_0.这个检验称为两样本 t 检验.拒绝域(8.2.10)的直观意义类似于(8.2.6).

需要说明的是,在上面的讨论中,我们假定了两个正态总体的方差相等.当然,这是一个迫不得已而强加上去的条件.因为,如果放弃这个假设,(8.2.8)就不再成立,我们也就无法使用简单易行的 t 检验.在实用中,只要我们有理由认为 σ_1^2 和 σ_2^2 相差不是太大,(8.2.8)式就近似成立,于是,相应的检验(8.2.10)式还是可行的.

例 8.2.3 假设有 A、B 两种药,试验者欲比较它们在服用 2 小时后血液中的含量是否一样.对药品 A,随机抽取 8 个病人,他们服药 2 小时后,测得血液中药的浓度(用适当的单位)为

$$1.23,\ 1.42,\ 1.41,\ 1.62,\ 1.55,\ 1.51,\ 1.60,\ 1.76.$$

对药品 B,随机抽取了 6 个病人,他们服药 2 小时后,测得血液中药的浓度为

$$1.76,\ 1.41,\ 1.87,\ 1.49,\ 1.67,\ 1.81.$$

假定这两组观测值服从具有公共方差的正态分布,试在显著性水平 $\alpha=0.10$ 下,检验病人血液中这两种药的浓度是否有显著不同?

解 记药品 A 和 B 的样本均值分别为 $\bar{X}$ 和 $\bar{Y}$,则 $\bar{X}=1.51$,$\bar{Y}=1.66$.它们的样本方差分别为 $S_1^2=0.03$,$S_2^2=0.21$,且由(8.2.7)式算得 $S=0.33$.又 $t_{m+n-2}\left(\frac{\alpha}{2}\right)=t_{8+6-2}(0.05)=t_{12}(0.05)=1.78$,故

$$|\bar{X}-\bar{Y}|=0.15<\sqrt{\frac{m+n}{mn}}St_{m+n-2}\left(\frac{\alpha}{2}\right)=\sqrt{\frac{8+6}{8\times 6}}\times 0.33\times 1.78=0.31.$$

于是,我们接受原假设.即认为病人血液中这两种药浓度无显著差异.

前面我们说明了拒绝域(8.2.6)和(8.2.10)的直观意义.它们都是用 $|\bar{X}-\bar{Y}|$ 的大小来判断原假设 $\mu_1=\mu_2$ 是否成立.当原假设成立时,$|\bar{X}-\bar{Y}|$ 应该跟零比较接近.所以当 $|\bar{X}-\bar{Y}|$ 大于某常数 c 时,我们就有理由认为原假设不成立而予以拒绝.基于这样的直观分析,我们可以处理如下更一般的假设检验问题.

(1) $H_0':\mu_1-\mu_2\geqslant 0 \longleftrightarrow H_1':\mu_1-\mu_2<0$;

(2) $H_0'': \mu_1 - \mu_2 \leqslant 0 \longleftrightarrow H_1'': \mu_1 - \mu_2 > 0$.

对问题(1)而言,一个适当的检验应当是,当 $\bar{X} - \bar{Y} \leqslant c$ 时,拒绝原假设 H_0',否则就接受 H_0'.

若 σ_1^2 和 σ_2^2 已知,根据(8.2.5),对给定的显著性水平 α,当

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} \leqslant - Z_\alpha$$

时,拒绝 H_0',否则就接受 H_0'. 于是 H_0' 的拒绝域为

$$\bar{X} - \bar{Y} \leqslant - Z_\alpha \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}. \tag{8.2.11}$$

完全类似的方法,可以得到 H_0'' 的拒绝域为

$$\bar{X} - \bar{Y} \geqslant Z_\alpha \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}. \tag{8.2.12}$$

当假定 σ_1^2 和 σ_2^2 未知但 $\sigma_1^2 = \sigma_2^2$ 时,用(8.2.9)代替(8.2.5),于是 H_0' 和 H_0'' 的拒绝域分别为

$$\bar{X} - \bar{Y} \leqslant -\sqrt{\frac{m+n}{mn}} S t_{m+n-2}(\alpha) \tag{8.2.13}$$

和

$$\bar{X} - \bar{Y} \geqslant \sqrt{\frac{m+n}{mn}} S t_{m+n-2}(\alpha). \tag{8.2.14}$$

三、成对数据的 t 检验

在上段讨论的用于两个正态总体均值的比较检验中,我们实际上是假设了来自这两个正态总体的样本是相互独立的. 但是,在实际中,有时候情况不总是这样. 可能这两个正态总体的样本是来自同一个总体上的重复测量,它们是成对出现的且是相关的. 例如,为了考察一种降血压药的效果,测试了 n 个高血压病人服药前后的血压分别为 $X_1, X_2, \cdots, X_n$ 和 $Y_1, Y_2, \cdots, Y_n$. 这里 (X_i, Y_i) 是第 i 个病人服药前和服药后的血压. 它们是有关系的,不会相互独立. 另一方面,$X_1, X_2, \cdots, X_n$ 是 n 个不同病人的血压,由于各人体质诸方面的条件不同,这 n 个观测值也不能看成来自同一个正态总体的样本. $Y_1, Y_2, \cdots, Y_n$ 也一样. 这样的数据称为成对数据. 对这样的数据在(二)中所讨论的检验方法就不适用. 但是,因为 X_i 和 Y_i 是在同一个人身上观测到的血压,所

以，$X_i - Y_i$ 就消除了人的体质诸方面的条件差异，仅剩下降血压药的效果．从而我们可以把 $d_i = X_i - Y_i, i = 1,2,\cdots,n$ 看成来自正态总体$N(\mu,\sigma^2)$的样本．其中 μ 就是降血压药的平均效果．降血压药是否有效，就归结为检验如下假设

$$H_0: \mu = 0 \longleftrightarrow H_1: \mu \neq 0.$$

因为 $d_1, d_2, \cdots, d_n$ 为来自正态总体$N(\mu,\sigma^2)$的随机样本，于是问题就变成了（一）款中 $\mu_0 = 0$ 的特殊情形，若记

$$\bar{d} = \sum_{i=1}^{n} d_i / n,$$

$$S_d^2 = \sum_{i=1}^{n} (d_i - \bar{d})^2 / (n-1),$$

那么根据(8.2.2)，我们就得到原假设 H_0 的显著性水平为 α 的检验的拒绝域为

$$|\bar{d}| \geqslant \frac{S_d}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right).$$

这个检验通常称为成对 t 检验．

例 8.2.4　为了检验 A、B 两种测定铁矿石含铁量的方法是否有明显差异，现用这两种方法测定了取自 12 个不同铁矿的矿石标本的含铁量(%)，结果列于表 8.2.1．问这两种测定方法是否有显著差异？取 $\alpha = 0.05$．

表 8.2.1　铁矿石含铁量(%)

标本号	方法 A	方法 B	d_i
1	38.25	38.27	-0.02
2	31.68	31.71	-0.03
3	26.24	26.22	+0.02
4	41.29	41.33	-0.04
5	44.81	44.80	+0.01
6	46.37	46.39	-0.02
7	35.42	35.46	-0.04
8	38.41	38.39	+0.02
9	42.68	42.72	-0.04
10	46.71	46.76	-0.05
11	29.20	29.18	+0.02
12	30.76	30.79	-0.03

解　将方法 A 和方法 B 的测定位分别记为 $X_1,X_2,\cdots,X_{12}$ 和 $Y_1,Y_2,\cdots,Y_{12}$. 由于这 12 个标本来自不同铁矿,因此,$X_1,X_2,\cdots,X_{12}$ 不能看成来自同一个总体的样本,$Y_1,Y_2,\cdots,Y_{12}$ 也一样. 故需用成对 t 检验. 记

$d_i = X_i - Y_i, i = 1,2,\cdots,12.\ \bar{d} = -0.0167, S_d^2 = 0.0007$. 查表得 $t_{n-1}\left(\dfrac{\alpha}{2}\right) = t_{11}(0.025) = 2.201$. 因为

$$\frac{S_d}{\sqrt{n}}t_{n-1}\left(\frac{\alpha}{2}\right) = \frac{\sqrt{0.0007}}{\sqrt{12}} \times 2.201 = 0.0168 > |\bar{d}| = 0.0167,$$

所以我们接受原假设,即认为两种测定方法无显著性差异.

§8.3　正态总体方差的检验

本节讨论关于正态总体方差的检验. 这里分一个正态总体方差的 χ^2 检验和两个正态总体方差比的 F 检验. 相对于正态总体均值的检验,方差检验的重要性要逊色得多,但也有一些应用,例如,机器所加工出的产品的尺寸服从正态分布. 这个正态分布的方差刻画了生产过程的稳定性. 方差越大,表示整个生产过程综合误差越大. 因此,我们需要知道方差是否超过了一个预定界限. 方差比的 F 检验主要用于上节讨论的两样本 t 检验中,关于两正态总体方差相等的假设是否合理.

一、单个正态总体方差的 χ^2 检验

设 $X_1,X_2,\cdots,X_n$ 为来自正态总体 $N(\mu,\sigma^2)$ 的样本,μ 和 σ^2 皆未知. 给定显著性水平 α,且设 σ_0^2 为给定的常数,我们要检验

$$H_0:\sigma^2 = \sigma_0^2 \longleftrightarrow H_1:\sigma^2 \neq \sigma_0^2.$$

我们知道样本方差 $S^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$ 是 σ^2 的一个无偏估计,于是,当原假设成立时,S^2 和 σ_0^2 应该比较接近,即比值 S^2/σ_0^2 应比较接近于 1. 因此,这个比值过大或过小都是我们拒绝原假设的理由. 根据基本定理(定理 6.4.1),在原假设下

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2, \tag{8.3.1}$$

取它为检验统计量,一个直观上合理的检验是当

$$\frac{(n-1)S^2}{\sigma_0^2} \leqslant c_1 \text{ 或 } \frac{(n-1)S^2}{\sigma_0^2} \geqslant c_2$$

时,拒绝原假设.这里 c_1 和 c_2 由显著性水平 α 来确定.为简单计,可取

$$c_1 = \chi_{n-1}^2\left(1-\frac{\alpha}{2}\right),\ c_2 = \chi_{n-1}^2\left(\frac{\alpha}{2}\right).$$

于是,检验的拒绝域为

$$\frac{(n-1)S^2}{\sigma_0^2} \leqslant \chi_{n-1}^2\left(1-\frac{\alpha}{2}\right) \text{ 或 } \frac{(n-1)S^2}{\sigma_0^2} \geqslant \chi_{n-1}^2\left(\frac{\alpha}{2}\right). \tag{8.3.2}$$

如果我们要检验假设

$$H_0:\sigma^2 \leqslant \sigma_0^2 \longleftrightarrow H_1:\sigma^2 > \sigma_0^2,$$

那么,当原假设成立时,$\sigma^2/\sigma_0^2 \leqslant 1$,因此,$S^2/\sigma_0^2$ 也倾向于比较小,根据(8.3.1),一个直观上合理的检验应当是,当

$$\frac{(n-1)S^2}{\sigma_0^2} \geqslant c$$

时,拒绝原假设.于是对给定的显著性水平 α,检验的拒绝域为

$$\frac{(n-1)S^2}{\sigma_0^2} \geqslant \chi_{n-1}^2(\alpha). \tag{8.3.3}$$

上面这两个检验都用到了 χ^2 分布,文献中把这类检验通称为 χ^2 检验.

例 8.3.1　某公司生产的发动机部件的直径服从正态分布.该公司称它的标准差 $\sigma = 0.048$ 厘米,现随机抽取 5 个部件,测得它们的直径为 1.32, 1.55, 1.36, 1.40, 1.44.取 $\alpha = 0.05$.问:(1) 我们能够认为该公司生产的发动机部件的直径的标准差确实为 $\sigma = 0.048$ 厘米吗? (2) 我们能否认为 $\sigma^2 \leqslant 0.048^2$?

解　(1) 本题要求在水平 $\alpha = 0.05$ 下检验假设

$$H_0:\sigma^2 = 0.048^2 \longleftrightarrow H_1:\sigma^2 \neq 0.048^2,$$

这里 $n = 5$, $\chi_{n-1}^2\left(\frac{\alpha}{2}\right) = \chi_4^2(0.025) = 11.143$, $\chi_{n-1}^2\left(1-\frac{\alpha}{2}\right) = \chi_4^2(0.975) = 0.$ 不能断

484.另一方面,计算得 $S^2 = 0.00778$.因为

$$\frac{(n-1)S^2}{\sigma_0^2} = \frac{(5-1)\times 0.00778}{0.048^2} = 13.51 > 11.143,$$

由(8.3.2)知,我们应该拒绝 H_0,即认为发动机部件的直径标准差不是 0.048

厘米.

(2) 本题要求在水平 $\alpha = 0.05$ 下检验假设

$$H_0:\sigma^2 \leqslant 0.048^2 \longleftrightarrow H_1:\sigma^2 > 0.048^2.$$

查表知 $\chi^2_{n-1}(\alpha) = \chi^2_4(0.05) = 9.49$. 因为

$$\frac{(n-1)S^2}{\sigma_0^2} = 13.51 > 9.49,$$

于是,由(8.3.3)知,我们应该拒绝原假设,即认为发动机原部件的直径标准差超过了 0.048.

二、两个正态总体方差比的 F 检验

设有两个正态总体 $N(\mu_1,\sigma_1^2)$ 和 $N(\mu_2,\sigma_2^2)$,我们欲检验

$$H_0:\sigma_1^2 = \sigma_2^2 \longleftrightarrow H_1:\sigma_1^2 \neq \sigma_2^2.$$

现在从这两个总体中分别抽取样本 $X_1,X_2,\cdots,X_m$ 和 $Y_1,Y_2,\cdots,Y_n$,它们是相互独立的. 分别以 S_1^2 和 S_2^2 记它们的样本方差. 直观上,S_1^2/S_2^2 是 σ_1^2/σ_2^2 的一个估计,当原假设成立时,后者等于 1,作为它们的估计,S_1^2/S_2^2 也应与 1 相差不远. 因此,一个直观上合理的拒绝域应为

$$\frac{S_1^2}{S_2^2} \leqslant c_1 \text{ 或 } \frac{S_1^2}{S_2^2} \geqslant c_2, \tag{8.3.4}$$

c_1 和 c_2 应根据显著性水平 α 和 S_1^2/S_2^2 的分布来确定. 由基本定理(定理 6.4.1)知

$\mu_1\cdot\mu_2$ 已知

$$\frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1},$$

$$\frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n-1},$$

它们相互独立,于是

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{m-1,n-1},$$

因而,当原假设成立时,从 $\sigma_1^2 = \sigma_2^2$ 可知

$$S_1^2/S_2^2 \sim F_{m-1,n-1}.$$

据此,我们可以选择显著性水平 α 下的拒绝域为

$$\frac{S_1^2}{S_2^2} \leqslant F_{m-1,n-1}\left(1-\frac{\alpha}{2}\right) \text{ 或 } \frac{S_1^2}{S_2^2} \geqslant F_{m-1,n-1}\left(\frac{\alpha}{2}\right). \tag{8.3.5}$$

通过类似的讨论,读者不难导出检验问题

$$H_0': \sigma_1^2 \leqslant \sigma_2^2 \longleftrightarrow H_1': \sigma_1^2 > \sigma_2^2.$$

显著性水平为 α 的检验的拒绝域为

$$\frac{S_1^2}{S_2^2} \geqslant F_{m-1,n-1}(\alpha). \tag{8.3.6}$$

上面这两个检验都用到了 F 分布,文献中把它们以及类似的检验通称为 F 检验.

例 8.3.2　甲、乙两厂生产同一种电阻,现从甲乙两厂的产品中分别随机抽取 12 个和 10 个样品,测得它们的电阻值后,计算出样本方差分别为 $S_1^2 = 1.40$, $S_2^2 = 4.38$. 假设电阻值服从正态分布,在显著性水平 $\alpha = 0.10$ 下,我们是否可以认为两厂生产的电阻阻值的方差:

(1) $\sigma_1^2 = \sigma_2^2$;　(2) $\sigma_1^2 \leqslant \sigma_2^2$.

解　(1) 该问题即检验假设:

$$H_0: \sigma_1^2 = \sigma_2^2 \longleftrightarrow H_1: \sigma_1^2 \neq \sigma_2^2.$$

因为 $m = 12, n = 10$,从(8.3.5)知,我们需要计算 $F_{11,9}(0.95)$,但一般 F 分布表中查不到这个值. 利用 F 分布的性质(见(6.4.1)式)有

$$F_{11,9}(0.95) = \frac{1}{F_{9,11}(0.05)} = \frac{1}{2.9} = 0.34,$$

而

$$\frac{S_1^2}{S_2^2} = \frac{1.40}{4.38} = 0.32 < 0.34 = F_{11,9}(0.95),$$

因此(8.3.5)的第一个不等式成立,所以,我们拒绝原假设,即认为两厂生产的电阻阻值的方差不同.

(2) 我们需要查 $F_{m-1,n-1}(\alpha) = F_{11,9}(0.10)$ 的值,但是在普通的 F 分布表中,查不到这个值. 于是我们用 $F_{10,9}(0.10)$ 和 $F_{12,9}(0.10)$ 的平均值作为它的近似,故有

$$F_{11,9}(0.10) = \frac{1}{2}[F_{10,9}(0.10) + F_{12,9}(0.10)]$$

$$= \frac{1}{2}(2.42 + 2.38) = 2.40.$$

但是，$S_1^2/S_2^2 = 0.34 < 2.40$，于是，我们接受原假设，即认为甲厂生产的电阻的阻值的方差（即波动性）较小.

§8.4　拟合优度检验

在前面的讨论中，我们总是假定总体分布形式是已知的. 例如，我们常说 $X_1, X_2, \cdots, X_n$ 为来自正态总体 $N(\mu, \sigma^2)$ 的样本. 在这里有一个重要问题，就是总体到底是不是正态分布 $N(\mu, \sigma^2)$，有时并不是很显然的，这就需要通过一定的检验. 本节所要讨论的拟合优度检验，就是为了这一目的而设计的. 一言以蔽之，拟合优度检验，就是检验观测到的一批数据是否服从某一特定的分布.

设 $F(x)$ 为一已知分布，现有样本 $X_1, X_2, \cdots, X_n$，但我们并不知道样本 $X_1, X_2, \cdots, X_n$ 的总体分布是什么. 现在试图检验

H_0：样本 $X_1, X_2, \cdots, X_n$ 的总体分布为 $F(x)$.　　(8.4.1)

对立假设就是这些样本的总体分布不是 $F(x)$. 如果 $F(x)$ 带未知参数 $\theta = (\theta_1, \cdots, \theta_r)'$，则应记为 $F(x, \theta)$，不妨设该分布是连续型的.

构造拟合优度检验的思想和步骤如下：

(1) 把 $(-\infty, \infty)$ 分割为 k 个区间：

$$-\infty = a_0 < a_1 < a_2 < \cdots < a_{k-1} < a_k = \infty,$$

记

$$I_1 = (a_0, a_1],\ I_2 = (a_1, a_2], \cdots, I_k = (a_{k-1}, a_k].$$

(2) 计算每个区间上的理论频数.

如果总体的分布为 $F(x, \theta)$，那么在区间 I_i 上，有理论概率

$$p_i(\theta) = F(a_i, \theta) - F(a_{i-1}, \theta), i = 1, 2, \cdots, k \qquad (8.4.2)$$

现在样本个数为 n，它们落在区间 I_i 上的应该有 $np_i(\theta)$ 个，这里 $np_i(\theta)$ 称为理论频数. 如果 θ 是未知的，用 $\hat{\theta}$ 表示它的极大似然估计，代入(8.4.2)，得到 $p_i(\hat{\theta})$，这时理论频数为 $np_i(\hat{\theta})$.

(3) 计算每个区间上的实际频数.

设 $X_1, X_2, \cdots, X_n$ 中有 f_i 个落在区间 I_i 上，称 f_i 为实际频数.

(4) 计算理论频数与实际频数的偏差平方和

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})}, \tag{8.4.3}$$

这里对每一项用 $np_i(\hat{\theta})$ 去除，其目的是缩小理论频数 $np_i(\hat{\theta})$ 比较大的那些项在和式中的影响力.

很明显，$np_i(\hat{\theta})$ 是从分布 $F(x,\theta)$ 算出来的区间 I_i 上的频数，而 f_i 是样本中落在 I_i 上的实际频数. 它们差异的大小就度量了这些样本与分布 $F(x,\theta)$ 拟合的程度. 统计量 χ^2 称为 Pearson 的拟合优度 χ^2 统计量，简称为 χ^2 统计量. 可以证明，若 H_0 成立，则当样本个数 $n \to \infty$ 时

$$\chi^2 \to \chi^2_{k-r-1}, \tag{8.4.4}$$

即 χ^2 统计量的分布收敛到自由度为 $k-r-1$ 的 χ^2 分布，这里 k 是划分的区间的个数，r 是被估计的参数的个数.

(5) 假设 H_0 的显著性水平为 α 的检验拒绝域为

$$\chi^2 \geqslant \chi^2_{k-r-1}(\alpha), \tag{8.4.5}$$

这个检验称为 χ^2 检验.

注 1 因为(8.4.4)是样本个数 $n \to \infty$ 时统计量 χ^2 的极限分布，所以在应用上要求 n 比较大，一般经验上认为 $n \geqslant 50$，并且每个 $np_i(\theta)$ 最好不小于 5. 如果 $(-\infty, \infty)$ 的初始划分不满足后一个条件时，则需将相邻区间合并，以满足这个要求.

(大)

注 2 χ^2 统计量可以直观地写为

$$\chi^2 = \sum_{i=1}^{k} \frac{(\text{经验频数} - \text{理论频数})^2}{\text{理论频数}}.$$

下面我们通过几个例子来说明 χ^2 统计量的应用.

例 8.4.1 为检验棉纱的拉力强度(单位：公斤) X 服从正态分布，从一批棉纱中随机抽取 300 条进行拉力试验，结果列在表 8.4.1 中，我们的问题是检验假设

$$H_0: \text{拉力强度 } X \sim N(\mu, \sigma^2) \quad (\alpha = 0.01).$$

(1) 将观测值 X_i 分成 13 组，这相当于 $a_0 = -\infty, a_1 = 0.64, a_2 = 0.78, \cdots, a_{12} = 2.18, a_{13} = \infty$ 但是这样分组后，前两组和最后两组的 $np_i(\theta)$ 比较小，于是我们把它们合并成为一个组(见表 8.4.2)

(应计算出)

表 8.4.1　棉纱拉力数据

i	x	n_i	i	x	n_i
1	0.5 ~ 0.64	1	8	1.48 ~ 1.62	53
2	0.64 ~ 0.78	2	9	1.62 ~ 1.76	25
3	0.78 ~ 0.92	9	10	1.76 ~ 1.90	19
4	0.92 ~ 1.06	25	11	1.90 ~ 2.04	16
5	1.06 ~ 1.20	37	12	2.04 ~ 2.18	3
6	1.20 ~ 1.34	53	13	2.18 ~ 2.38	1
7	1.34 ~ 1.48	56			

表 8.4.2　棉纱拉力数据的分组

区间序号	区间	f_i	$p_i(\hat{\theta})$	$np_i(\hat{\theta})$	$f_i - np_i(\hat{\theta})$
1	⩽0.78 或 >2.04	7	0.0156	4.68	2.32
2	0.78 ~ 0.92	9	0.0223	6.69	2.31
3	0.92 ~ 1.06	25	0.0584	17.52	7.48
4	1.06 ~ 1.20	37	0.1205	36.15	0.85
5	1.20 ~ 1.34	53	0.1846	55.38	−2.38
6	1.34 ~ 1.48	56	0.2128	63.84	−7.84
7	1.48 ~ 1.62	53	0.1846	55.38	−2.38
8	1.62 ~ 1.76	25	0.1205	36.15	−11.15
9	1.76 ~ 1.90	19	0.0584	17.52	1.48
10	1.90 ~ 2.04	16	0.0223	6.69	9.31

(2) 计算每个区间上的理论频数. 这里 $F(x,\theta)$ 就是正态分布 $N(\mu,\sigma^2)$ 的分布函数，含有两个未知参数 μ 和 σ^2，分别用它们的极大似然估计 $\mu^* = \overline{X}$ 和 $\sigma^{*2} = \sum_{i=1}^{n}(X_i - \overline{X})^2/n$ 来代替. 关于 $\overline{X}$ 的计算作如下说明. 因为表8.4.1中每个区间都很狭窄，我们可以认为每个区间内 X_i 都取这个区间的中点，然后将每个区间的中点值乘以该区间的样本数，将这些值相加再除以总样本数就得到样本均值 $\overline{X}$，具体计算结果：$\mu^* = 1.41, \sigma^{*2} = 0.26^2$. 对于服从 $N(1.41, 0.26^2)$ 的随机变量 Y，计算它在上面每个区间上的概率 $p(\hat{\theta})$，例如，

$$
\begin{aligned}
p_1(\hat{\theta}) &= P\{Y \leqslant 0.78\} + P\{Y > 2.04\} \\
&= P\left\{\frac{Y - 1.41}{0.26} \leqslant -2.42\right\} + P\left\{\frac{Y - 1.41}{0.26} > 2.42\right\} \\
&= 0.0156,
\end{aligned}
$$

$$
\begin{aligned}
p_2(\hat{\theta}) &= P\{0.78 < Y \leqslant 0.92\} \\
&= P\left\{-2.42 < \frac{Y-1.41}{0.26} \leqslant -1.88\right\} \\
&= 0.0223.
\end{aligned}
$$

(3) 计算 $X_1, X_2, \cdots, X_{300}$ 中落在每个区间上的实际频数 f_i，如表 8.4.2 所列.

(4) 计算统计量的值：

$$
\chi^2 = \sum_{k=1}^{10} \frac{(f_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})} = 22.07.
$$

因为 $k = 10, r = 2$，所以，χ^2 的自由度为 $10-2-1=7$，查表得$\chi_7^2(0.01) = 18.48 < \chi^2 = 22.07$. 于是，我们拒绝原假设，即认为棉纱拉力强度不服从正态分布.

χ^2 检验的一个著名应用例子是孟德尔豌豆试验. 奥地利生物学家孟德尔在 1865 年发表的论文，事实上提出了基因学说，奠定了现代遗传学的基础. 他的这项伟大发现的过程有力地证明了统计方法在科学研究中的作用. 因此，我们有必要在这里将这一情况介绍给读者.

孟德尔在关于遗传问题的研究中，用豌豆做试验. 豌豆有黄、绿两种颜色，在对它们进行两代杂交之后，发现一部分杂交豌豆呈黄色，另一部分呈绿色. 其数目的比例大致是 3 : 1. 孟德尔把他的试验重复了多次，每次都得到类似结果. 这只是一个表面上的统计规律. 但它启发孟德尔去发展一种理论，以解释这种现象. 他大胆地假定存在一种实体，即现在我们称为"基因"的东西，决定了豌豆的颜色. 这基因有黄绿两个状态，一共有四种组合：

(黄，黄)，(黄，绿)，(绿，黄)，(绿，绿)

孟德尔认为，前三种配合使豆子呈黄色，而第四种配合使豆子呈绿色. 从古典概率的观点看，黄色豆子出现的概率为 3/4，而绿色豆子出现的概率为 1/4. 这就解释了黄绿颜色豆子之比为什么总是接近 3 : 1 这个观察结果. 孟德尔这个发现的深远意义是他开辟了遗传学研究的新纪元. 下面的例子就是用 χ^2 检验来检验孟德尔提出黄绿颜色豌豆数目之比为 3 : 1 的论断.

例 8.4.2 孟德尔豌豆试验中，发现黄色豌豆为 25 个，绿色豆 11 个，试在 $\alpha = 0.05$ 下，检验 3 : 1 这个比例.

解 定义随机变量 X

$$
X = \begin{cases} 1, & \text{若豆是黄色}, \\ 0, & \text{若豆是绿色}. \end{cases}
$$

记 $p_1 = P(X = 1), p_2 = P(X = 0)$,我们要检验假设

$$H_0: p_1 = \frac{3}{4}, p_2 = \frac{1}{4}.$$

(1) 将 $(-\infty, +\infty)$ 分成两个区间 $I_1 = (\frac{1}{2}, \infty), I_2 = (-\infty, \frac{1}{2}]$,

(2) 计算每个区间上的理论频数,这里 $n = 25 + 11 = 36$,不存在要估计的未知参数,故 $np_1 = 36 \times \frac{3}{4} = 27, np_2 = 36 \times \frac{1}{4} = 9$.

(3) 实际频数为　$f_1 = 25, f_2 = 11$.

(4)

$$\chi^2 = \frac{(25-27)^2}{27} + \frac{(11-9)^2}{9}$$

$$= 0.148 + 0.444 = 0.592.$$

查表得 $\chi^2_{k-1}(0.05) = \chi^2_1(0.05) = 3.84 > 0.592 = \chi^2$. 所以,我们接受原假设,即认为黄绿色豌豆数目之比为 3 : 1.

§8.5　独立性检验

许多社会调查数据往往可以总结成一种表格形式. 这种表格在统计学上称为列联表. 例如,我们调查了 520 个人,其中 136 人患有高血压,另外的 384 人血压正常. 另一方面在患高血压的 136 人中,有 48 人有冠心病,其余的 88 人无此病. 在无高血压的 384 人中,有 36 人患有冠心病,将这些数据列成表格 8.5.1,我们希望考察高血压和冠心病两者之间是否有关系. 即欲考察列联表中的两个因素是否独立. 我们可以用上节介绍的 χ^2 检验,来解决这个问题.

表 8.5.1　2 × 2 列联表

	患高血压	无高血压	总　计
患冠心病	48	36	84
无冠心病	88	348	436
总　计	136	384	520

本例中,我们考察两个因素,一是"高血压",它有两个状态,统计学上称为水平:"患高血压" 和"无高血压". 另一个因素是"冠心病",它也有两个水平:"患冠心病" 和"无冠心病". 这种表称为 2×2 列联表. 一般地,设有 A 和 B 两个

因素，各有 a 和 b 个水平，问题是要检验 A 和 B 两个因素相互独立. 随机观察了 n 个对象，其中有 n_{ij} 个对象的因素 A 和 B 分别处在水平 i 和 j. 其中

表 8.5.2 $a\times b$ 列联表

A \ B	1	2	…	j	…	b	行和
1	n_{11}	n_{12}	…	n_{1j}	…	n_{1b}	$n_{1\cdot}$
2	n_{21}	n_{22}	…	n_{2j}	…	n_{2b}	$n_{2\cdot}$
⋮	⋮	⋮	…	⋮	…	⋮	⋮
i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{ib}	$n_{i\cdot}$
⋮	⋮	⋮	…	⋮	…	⋮	⋮
a	n_{a1}	n_{a2}	…	n_{aj}	…	n_{ab}	$n_{a\cdot}$
列和	$n_{\cdot 1}$	$n_{\cdot 2}$	…	$n_{\cdot j}$	…	$n_{\cdot b}$	n

$$n_{i\cdot} = \sum_j n_{ij}\,,\quad n_{\cdot j} = \sum_i n_{ij} \tag{8.5.1}$$

分别表示表 8.5.2 中的第 i 行之和与第 j 列之和，即 $n_{i\cdot}$ 为 n 个对象中其因素 A 在水平 i 的个数，而 $n_{\cdot j}$ 为 n 个对象中其因素 B 在水平 j 的个数. 如果记

$$p_{ij} = P(\text{因素 } A \text{ 和 } B \text{ 分别处在水平 } i \text{ 和 } j),$$

$$p_{i\cdot} = P(\text{因素 } A \text{ 处在水平 } i),\quad i = 1,\cdots,a,$$

$$p_{\cdot j} = P(\text{因素 } B \text{ 处于水平 } j),\quad j = 1,\cdots,b.$$

我们可以把 p_{ij} 理解为对象的因素 A 和 B 分别处在水平 i 和 j 的理论概率，余类推. 我们要检验的假设是 A 与 B 相互独立，它可以表为如下独立性假设

$$H_0: p_{ij} = p_{i\cdot}\, p_{\cdot j}\,,\ i = 1,\cdots,a,\quad j = 1,\cdots,b. \tag{8.5.2}$$

这个检验问题，可以通过上节的 χ^2 检验来完成. 我们可以把表 8.5.2 中的 ab 个格子，看成 ab 个区间，n_{ij} 就是每个格子的实际频数. 根据独立性假设 (8.5.2)，$np_{ij} = np_{i\cdot}\, p_{\cdot j}$ 就是理论频数，于是在原假设下

$$\chi^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(n_{ij}-np_{ij})^2}{np_{ij}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(n_{ij}-np_{i\cdot}\, p_{\cdot j})^2}{np_{i\cdot}\, p_{\cdot j}}, \tag{8.5.3}$$

这就刻画了实际数据与理论假设 H_0 拟合的程度. 当 n 很大时，这个 χ^2 统计量近似地服从 χ^2 分布. 但这里还有两个问题：$p_{i\cdot}$ 和 $p_{\cdot j}$ 的估计和 χ^2 的自由度的计算.

根据用“频率估计概率”的原则，$p_{i\cdot}$ 和 $p_{\cdot j}$ 一个直观上很自然的估计分别是

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}, \quad i = 1,\cdots,a, \tag{8.5.4}$$

$$\hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}, \quad j = 1,\cdots,b. \tag{8.5.5}$$

从形式看，我们估计了 a 个 $p_{i\cdot}$ 和 b 个 $p_{\cdot j}$，但是注意到 $\sum\limits_{i=1}^{a} p_{i\cdot} = 1$，$\sum\limits_{j=1}^{b} p_{\cdot j} = 1$，所以，我们实际上估计了 $a + b - 2$ 个参数. 于是，χ^2 分布的自由度为

$$ab - (a + b - 2) - 1 = (a - 1)(b - 1).$$

将(8.5.4)和(8.5.5)代入(8.5.3)得

$$\chi^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(nn_{ij} - n_{i\cdot}n_{\cdot j})^2}{nn_{i\cdot}n_{\cdot j}}. \tag{8.5.6}$$

当 n 很大时，它近似地服从 $\chi^2_{(a-1)(b-1)}$. 于是，对给定的显著性水平 α，当 $\chi^2 \geqslant \chi^2_{(a-1)(b-1)}(\alpha)$ 时，拒绝原假设，即认为因素 A 与 B 不独立.

例 8.5.1 对表 8.5.1 中的数据，检验假设：高血压与冠心病无关系($\alpha = 0.05$).

解 因为 $n = 520, n_{11} = 48, n_{12} = 36, n_{21} = 88, n_{22} = 348, n_{1\cdot} = 84, n_{2\cdot} = 436, n_{\cdot 1} = 136, n_{\cdot 2} = 384$，于是

$$\chi^2 = \frac{(520 \times 48 - 84 \times 136)^2}{520 \times 84 \times 136} + \frac{(520 \times 36 - 84 \times 384)^2}{520 \times 84 \times 384}$$
$$+ \frac{(520 \times 88 - 436 \times 136)^2}{520 \times 436 \times 136} + \frac{(520 \times 348 - 436 \times 384)^2}{520 \times 436 \times 384}$$
$$= 30.84 + 10.92 + 5.94 + 2.10 = 49.8.$$

查 χ^2 分布表，$\chi^2_{(a-1)(b-1)}(\alpha) = \chi^2_{(2-1)(2-1)}(0.05) = 3.84 < \chi^2 = 49.8$. 因此，我们拒绝原假设，即认为冠心病和高血压有密切关系.

例 8.5.2 1936 年瑞典在研究家庭小孩个数与其收入关系时，调查了 25263 个家庭，其数据总结成 5×4 列联表(收入以千瑞典克朗为单位)，见表 8.5.3. 我们要来检验的假设是：

H_0：每个家庭小孩数的多少与它们的收入无关($\alpha = 0.001$).

因为 $a = 5, b = 4$，所以这是 5×4 列联表，统计量 χ^2 中有 20 项. 先算(8.5.6)

中的对应于 $n_{11} = 2161$ 的那项

表 8.5.3　家庭收入与小孩个数数据

小孩数 \ 收入	0 ~ 1	1 ~ 2	2 ~ 3	3 以上	合计
0	2161	3577	2184	1636	9558
1	2755	5081	2222	1052	11110
2	936	1753	640	306	3635
3	225	419	96	38	778
4	39	98	31	14	182
合计	6116	10928	5173	3046	25263

$$\frac{(25263 \times 2161 - 9558 \times 6116)^2}{25263 \times 9558 \times 6116} = 10.11,$$

对应于 $n_{12} = 3577$ 的那项为

$$\frac{(25263 \times 3577 - 9558 \times 10928)^2}{25263 \times 9558 \times 10928} = 75.17.$$

我们先不继续算下去，χ^2 统计量的值一定大于这两项之和 85.28，此处自由度 $(5-1)(4-1) = 12$，查 χ^2 分布表得 $\chi^2_{12}(0.001) = 32.909$，$\chi^2 > 85.28 > \chi^2_{12}(0.001) = 32.909$. 因此我们应该拒绝原假设. 这个结果表明，即使在显著性水平 $\alpha = 0.001$ 下，我们仍认为小孩数与家庭收入有密切关系. 从表上可以看出一个基本趋势，家庭收入愈低，小孩个数愈多. (不[illegible])

习　题　八

8.1　某油品公司的桶装润滑油标定重量为 10 公斤. 商品检验部门从市场上随机抽取 10 桶，称得它们的重量分别是 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3 和 9.8(公斤). 假设每桶油实际重量服从正态分布. 试在显著性水平 $\alpha = 0.01$ 下，检验该公司的桶装润滑油重量是否确为 10 公斤？

8.2　假设香烟中尼古丁含量服从正态分布，现从某牌香烟中随机抽取 20 支，其尼古丁含量的平均值 $\bar{X} = 18.6$ 毫克，样本标准差 $S = 2.4$ 毫克. 取显著性水平 $\alpha = 0.01$，我们能否接受“该种香烟的尼古丁含量的均值 $\mu = 18$ 毫克”的断言？

8.3　(1) 考虑正态总体 $N(\mu, \sigma^2)$ 和假设检验问题

$$H_0: \mu \leqslant \mu_0 \longleftrightarrow H_1: \mu > \mu_0.$$

证明，当 σ^2 已知时，则拒绝域为

$$\bar{X} - \mu_0 \geqslant \frac{\sigma}{\sqrt{n}} Z_\alpha$$

的检验的显著性水平为 α. 若 σ^2 未知,则拒绝域为

$$\bar{X}-\mu_0 \geqslant \frac{S}{\sqrt{n}} t_{n-1}(\alpha)$$

的检验的显著性水平为 α.

(2) 在习题 8.2 中,对 $\sigma = 2.4$ 毫克和 $S = 2.4$ 毫克两种情况,我们能否接受"该牌的香烟尼古丁含量不超过 17.5 毫克"的断言?

8.4　设某厂生产的产品尺寸服从正态分布 $N(\mu,\sigma^2)$,规定标准尺寸为 120 毫米. 现从该厂抽得 5 件产品,测量其尺寸分别为

119, 120, 119.2, 119.7, 119.6.

试判断产品是否符合规定要求,即检验假设 $H_0:\mu = 120 \longleftrightarrow \mu \neq 120$. (显著性水平 $\alpha = 0.05$).

8.5　设甲、乙两煤矿所产的煤中含煤粉率分别为 $N(\mu_1, 7.5)$ 和 $N(\mu_2, 2.6)$. 为检验这两个煤矿的煤含煤粉率有无明显差异,从两矿中取样若干份,测试结果如下:

甲矿(%):24.3, 20.8, 23.7, 21.3, 17.4,

乙矿(%):18.2, 16.9, 20.2, 16.7.

试在显著性水平 $\alpha = 0.05$ 下,检验"含煤粉率无差异"这个假设.

8.6　比较 A、B 两种小麦品种蛋白质含量. 随机抽取 A 种小麦 10 个样品,测得 $\bar{X} = 14.3$, $S_1^2 = 1.62$. 随机抽取 B 种小麦 5 个样品,测得 $\bar{Y} = 11.7$, $S_2^2 = 0.14$. 假定这两种小麦蛋白质含量都服从正态分布,且具有相同方差,试在 $\alpha = 0.01$ 水平下,检验两种小麦的蛋白质含量有无差异?

8.7　由于存在声音反射的原因,人们在讲英语时在辅音识别上会遇到麻烦. 有人随机选取了 10 个以英语为母语的人(记为 A 组) 和 10 个以英语为外国语的(记为 B 组) 的人,进行了试验,下面记录了他们正确反应的比例(%).

A组:93, 85, 89, 81, 88, 88, 89, 85, 85, 87,

B组:76, 84, 78, 73, 78, 76, 70, 82, 79, 77.

假定这些数据都来自正态总体,且具有公共方差,试在 $\alpha = 0.05$ 下,检验这两组的反应是否有显著差异?

8.8　某厂生产的瓶装纯净水要求标准差 $\sigma = 0.02$ 升,现在从超级市场上随机抽取 20 瓶这样的纯净水,发现它们所装水量的样本标准差 $S = 0.03$ 升. 假定瓶装纯净水装水量服从正态分布,试问在显著性水平 $\alpha = 0.05$ 下,我们能否认为它们达到了标准差 $\sigma = 0.02$ 升的要求?

8.9　试写出检验(8.3.6) 的推导过程.

8.10　试对习题 8.7 的数据,检验假设

$$H_0:\sigma_1^2 = \sigma_2^2 \longleftrightarrow H_1:\sigma_1^2 \neq \sigma_2^2.$$

8.11　某种导线要求电阻标准差不超过 0.05 欧,今在生产的一批导线中随机抽取 9 根,测量后算得 $S = 0.07$ 欧. 设电阻测量值服从正态分布,问在 $\alpha = 0.05$ 下,能否认为这批导线的电阻值满足原来的要求?

8.12　孟德尔豌豆试验中，有一次观测到黄色和绿色豆子的数目分别为 70 和 27，试在显著性水平 $\alpha = 0.05$ 下，检验“黄色和绿色豆子的数目为 3 : 1”的理论.

8.13　在一个复杂试验中，孟德尔同时考虑豌豆的颜色和形状，一共有四种组合：(黄，圆)，(黄，非圆)，(绿，圆)，(绿，非圆). 按孟德尔理论这四类应有 9 : 3 : 3 : 1 的比例. 在一次观察中，他发现这四类观测到的数目分别为 315，101，108 和 32，试在 $\alpha = 0.05$ 下，检验“9 : 3 : 3 : 1”这个理论.

8.14　某汽车修理公司想知道每天送来修理的车数是否服从泊松分布. 下表给出了该公司 250 天的送修车数：

送修车数	0	1	2	3	4	5	6	7	8	9	10
送这么多车的天数	2	8	21	31	44	48	39	22	17	13	5

试在 $\alpha = 0.05$ 下，检验原假设 H_0：一天内送修车数服从泊松分布 $P(\lambda)$.

8.15　为检验一颗骰子的均匀性，对这颗骰子投掷 60 次，观察到出现 1，2，…，6 点的次数分别为 7，6，12，14，5，16. 试在 $\alpha = 0.05$ 下，检验原假设：这颗骰子是均匀的，即每个点出现的概率均为 1/6.

8.16　某工厂三种配方生产出来的产品质量如下表：

质量状况 \ 配方	1	2	3
合格品	63	47	65
废　品	16	7	3

试分别在 $\alpha = 0.05$ 和 $\alpha = 0.01$ 下，检验原假设：三种配方生产出的产品质量无明显差异.