5.14 某果园有 1 000 株果树，在采摘前欲估计果园的总产量，随机抽选了 10 株，产量分别为 161,68,45,102,38,87,100,92,76,90 公斤。假设果树的产量服从正态分布，试以 95% 的置信水平估计该果园的总产量，若要求置信区间缩小一半，应抽多少株果树作样本？

第六章 假设检验

假设检验和参数估计一样,是统计推断的一种形式。假设检验是用样本的观测结果对总体参数的取值或总体行为所做的预先假定进行验证,从而作出接受或是拒绝原来假设的选择。本章介绍假设检验的基本概念,以及假设检验的一般步骤,然后介绍各种常用的参数检验和非参数检验方法。

第一节 假设检验的基本概念

假设检验是统计推断的另一种形式,和估计一样要通过样本取得数据,对总体作出推断。估计是通过样本的观察结果来推断未知总体参数的取值范围和作出结论的可靠程度,而假设检验是预先对总体参数的取值作出假定,然后用样本数据来验证,作出接受还是拒绝原来假设的结论。当然由于样本的随机性,这种推断也同样有一定风险。先提出假设,后加以论证,再决定取舍是科学研究中常用的方法之一。统计的假设检验可以作为帮助研究人员和管理人员决策的一种手段。例如工业产品的质量管理就是应用了这一方法。质量控制图假设在正常生产的情况下产品的某一指标是围绕其平均值 μ 的上下微小变动,然后每隔一段时间抽验一定的产品样本,如果符合假设的要求,就视为正常,继续生产。如样本数据出现了变化,过高或过低,与原假设不符,那就要停止生产,检查原因,以避免生产次品。从本章后面的内容还可以看到产品的抽样验收等都是应用了假设检验的原理。

一、假设检验中的一些基本概念

(一) 原假设(null hypothesis)和备择假设(alternative hypothesis)

在假设检验的一开始,首先要提出一个假设,就称作原假设,又称零假设或虚拟假设等,通常用 H_0 表示。比如,在质量管理中假设在正常的情况下,零件的平均长度应是 2 厘米,就建立 $H_0: \mu = 2$ 厘米。在提出原假设的同时,还要制定另一个假设称做备择假设。原假设是待检验的假设,备择假设则是原假设被拒绝后替换的假设。因为对于任何一个假设检验问题所有可能的结果都应包含在两个假设之内,非此即彼。如上面质量管理的例子中,零件的平均长度要么等于 2 厘米,要么不等于 2 厘米,备择假设通常用 H_1 表示,因此可以建立 $H_1: \mu \neq 2$ 厘米。

(二) 检验统计量

对原假设检验时必然要根据样本的数据来判断。对样本数据进行加工并用来判断是否接受原假设的统计量称做检验统计量,如上面列举的原假设 $H_0: \mu = \mu_0$(μ_0 为一已知数值),那么样本均值 $\bar{X}$ 就可以作为检验统计量,有时为了方便还将样本均值进一步加工,如样本均值服从正态分布时将样本均值标准化,$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ 作为检验统计量,简称 Z 统计量,或者根据条件用 $t = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$,称作 t 统计量。统计量的选择要根据研究的参数及其估计量的分布、抽样的方式、总体方差是否已知等多种因素来确定。

(三) 接受域和拒绝域

假设检验根据检验统计量的具体结果来判别是否接受 H_0,因此在假设 H_0 为真的情况下将抽样所有可能结果组成的样本空间划分为两部分,一部分是原假设为真时允许范围内的变动,应该接受原假设,因此称作接受域;另一部分是超出了一定的界限,当原假设为真时只有很小的概率出现,因而当统计量的结果落入这一

区域便应拒绝原假设，这一区域便称作拒绝域。接受域和拒绝域之间的分割点通常称作临界值。为直观地说明接受域和拒绝域，见图 6.1 所示。

(四) 显著性水平

假设检验的基本原理是根据小概率原理。所谓小概率原理是指发生概率很小的事件在一次试验中几乎是不可能发生的，根据这一原理就可以作出是否接受原假设的决定。例如有一批产品共 1 000 个，生产者声称只有一个是次品，那么随机抽取 1 个作检验时，通常不会抽到次品，因为抽中次品是千分之一的小概率，但如果在一次抽取中抽中了次品，显然就有理由怀疑生产者的声称，认为 1 000 个中只有一个次品的说法是假的。在假设检验中也是一样。我们确定了原假设为真时的可能范围为接受域，而落入拒绝域是个小概率事件，一旦落入拒绝域，就要拒绝原假设而接受备择假设。那么应该确定多大的范围算作小概率呢？这要根据不同的研究对象来确定，有的选择 0.05，有的选择 0.01，通常用 α 表示，它说明用多大的小概率来检验原假设，显然 α 愈小愈不容易推翻原假设，而一旦拒绝原假设，原假设为真的可能性就越小。因此检验时通常首先要规定显著性水平 α。

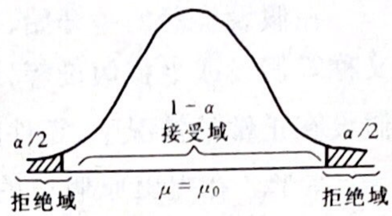
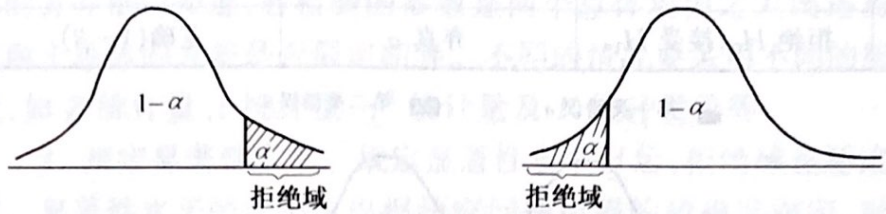


图 6.1　假设检验中 $\mu=\mu_0$ 时的接受域和拒绝域

(五) 双侧检验与单侧检验

假设检验根据实际的需要可以分为双侧检验和单侧检验，单侧检验又进一步分为左侧检验和右侧检验。双侧检验所针对的问题是指一些客体的指标过大和过小都不符合要求，因此都需要加以检验，这时检验的拒绝域就位于图形的两侧，如图 6.1。当显著性水平为 α 时，两边的拒绝域发生的概率各为 $\alpha/2$。但是在实际问题中有些现象的指标则要求愈低愈好，但不能超过某一标准，例如次品率。当超过这一临界点就要拒绝原假设，这就是单侧的假设检验，其拒绝域在图形的右侧，称作右侧检验。另外一些现象的

指标值则是愈高愈好而不能低于某值,如灯管的使用寿命,药物的有效成分等。当低于某一临界值就要拒绝原假设,这时拒绝域在图形的左侧,称作左侧检验,现用图 6.2 表示。



(a)右侧检验　　(b)左侧检验

图 6.2　假设检验中的单侧检验

(六) 假设检验中的两类错误

从假设检验的原理与规则可以看到,它是根据小概率原理来判断的,因此有可能会判断错误,因为在原假设为真的情况下,很可能有些样本统计量的估计值会落入小概率的拒绝域内而按决策规则加以拒绝。另外在原假设非真的情况下也有可能有一些统计量的估计值落入接受域的范围之内而接受原假设。因此可以把这些情况归结为两类错误。第一类错误是原假设 H_0 为真而却被拒绝的错误,它是犯了弃真的错误,犯错误的概率就是 α,所以也叫 α 错误或第一类错误。第二类错误是指原假设 H_0 为非真而却予以接受的错误,这是一种取伪的错误,这种错误发生的概率通常用 β 表示,故也称 β 错误或第二类错误。在生产者将产品售给消费者时,通常要进行产品的质量检验,原假设总是产品是合格的,但检验时生产者总是耽心把合格品检验为不合格品,也就是第一类错误 α,所以 α 也称为生产者风险。而在消费者一方却恐怕把不合格品检验不出来当作合格品而接受,因而第二类错误 β 也称为消费者风险。通常的假设检验只规定第一类错误 α,即显著性水平,而不考虑第二类错误 β,并称这样的检验为显著性检验。要设计一个好的检验方案应既要限制 α 错误,又要限制 β 错误。关于两类错误的关系可以用表 6.1 加以归纳和图 6.3 加以显示。

表 6.1　假设检验的可能结果

		H_0 为真	H_0 为假
决策	接受 H_0	正确$(1-\alpha)$	取伪 β
	拒绝 H_0,接受 H_1	弃真 α	正确$(1-\beta)$

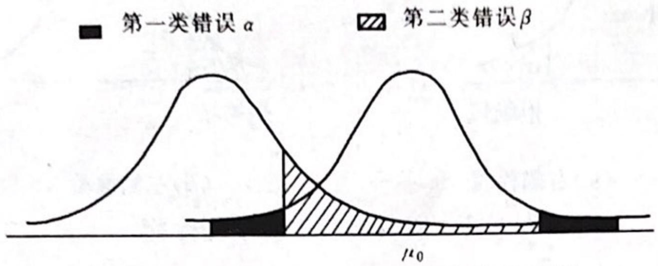


图 6.3　$H_0:\mu=\mu_0$ 时两类错误示意图

二、假设检验的一般步骤

1. 根据研究问题的需要建立原假设 H_0 和备择假设 H_1。由于假设检验是利用小概率原理来否定原假设,因此原假设总是与等号在一起。以均值的假设检验为例,不外乎三种情况:(1)$H_0:\mu=\mu_0$;$H_1:\mu\neq\mu_0$　(2)$H_0:\mu\geqslant\mu_0$;$H_1:\mu<\mu_0$　(3)$H_0:\mu\leqslant\mu_0$;$H_1:\mu>\mu_0$。其中(1)为双侧假设检验;(2)为左侧假设检验;(3)为右侧假设检验。采用哪一种检验要视研究的目的而定,尤其是在左侧检验和右侧检验时,放在 H_0 或 H_1 的不同位置往往会得出相反的结论。假设检验中所用的推断方法类似于数学中的反证法,把所希望证明的假设常常作为备择假设。因为放在 H_0 和 H_1 在接受时可能犯错误的概率是不同的,放在 H_0 接受时可能犯的错误是取伪错误,因此是 β,通常不容易知道。而放在 H_1 接受时可能犯错误的概率就是弃真错误的概率 α,是可以明确知道的。另外需要说明的是当接受 H_0 时,并不意味着 H_0 一定正确,而确切地说应该是根据样本数据在显著性水平为 α 的情况下尚不能推翻原假设。

2. 找出检验统计量及其分布。在建立好假设以后要确定 H_0 还是拒绝 H_0 都是根据检验统计量的具体结果落入接受域还是拒

绝域而定。这就要确定什么是检验统计量及该统计量服从什么分布。确定检验统计量及其分布(包括其数学期望和方差)是由许多因素决定的。如检验的是什么参数,总体的分布形式是否已知,总体的方差是否知道,若检验的参数是两个总体均值之差则还需知道两个总体的方差是否假定相等。不同的情况要采用不同的统计量,如 Z 统计量、t 统计量、χ^2 统计量及 F 统计量等等。

3. 规定显著性水平。规定显著性水平以后,拒绝域也随之而定。显著性水平的大小应根据研究问题所需的精确度而定,对于接受备择假设而言,如果要求结论比较精确,显著性水平 α 应该小一些,反之,要求不太精确,α 可稍大一些,可取 0.05 或 0.1。

但是,读者应当注意的是,显著性水平的大小有时会影响假设检验的结果。例如对同一个问题,当显著性水平 $\alpha=0.1$ 时拒绝了原假设,当 $\alpha=0.01$ 时就可能接受原假设。这一点可以很容易地从图 6.4 中看出来。

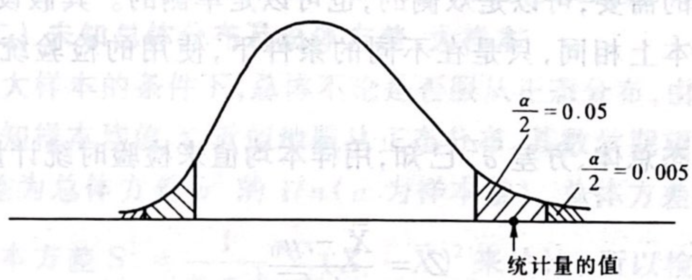


图 6.4　显著性水平的大小影响假设检验的结果

4. 确定决策规则。在规定了显著性水平与选择好检验的统计量后,就可以根据原假设与对立假设的设置情况找出接受域与拒绝域的临界点。例如统计量若服从正态分布,标准化后的统计量为 Z。当 $\alpha=0.05$ 双侧检验,其临界值为 ±1.96。当统计量 $|Z|>1.96$时就应拒绝 H_0,接受 H_1。不同的检验在规则上略有不同,将在后面的介绍中分别说明。

5. 根据观察所得到的数据进行计算并作出决策。

当然以上的各个步骤有些有先后关系,如先有了假设,确定了

检验统计量,才能有决策的规则,但显著性水平的规定可以先后灵活一些。

第二节　参数的假设检验

一、单个总体的均值检验

(一) 已知总体为正态分布且方差 σ^2 已知

这一小节要讨论的是检验某一客观现象已知服从正态分布,方差为 σ^2,但均值未知,现从中抽取一个样本欲检验其总体均值 μ 是否等于某个均值 μ_0,如质量管理中的均值控制图就是这种检验的应用。另一种情况是检验某一样本是否抽自一个均值为 μ_0,方差为 σ^2 的总体。因此通常是检验 $H_0: \mu = \mu_0$,其备择假设根据研究问题的需要,可以是双侧的,也可以是单侧的。其假设检验的步骤也基本上相同,只是在不同的条件下,使用的检验统计量不同。

当正态总体、方差 σ^2 已知,用样本均值来检验时统计量为:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \tag{6.1}$$

【例 6.1】　一个生产宇航飞行器的工厂需要经常购置一种耐高温的零件,要求抗热的平均温度是 1 250 ℃,在过去,供货者提供的产品都符合要求,并从大量的数据获知零件抗热的标准差是 150 ℃,在最近的一批进货中随机测试了 100 个零件,其平均的抗热为 1 200℃,能否接受这批产品?工厂希望对实际产品符合要求而错误地加以拒绝的风险为 0.05(即 $\alpha = 0.05$)。

解:检验的步骤如下:

(1) 建立假设。由于检验的目的是希望产品零件抗热的均值高于 1 250℃,而把低于 1 250℃ 的加以拒绝,因此是一个单侧的假

设检验问题。

$$H_0: \mu \geqslant 1\,250℃ \qquad H_1: \mu < 1\,250℃$$

(2) 这个检验中适当的检验统计量是:

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

(3) 根据工厂的要求,显著性水平 $\alpha = 0.05$,在这里是指当 $\mu = 1\,250$时而被拒绝的概率为 α。

(4) 根据单侧检验 $\alpha = 0.05$ 时,Z 统计量拒绝域的临界值为 $-z_\alpha = -z_{0.05} = -1.645$。

(5) 计算统计量的数值

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{1\,200 - 1\,250}{150 / \sqrt{100}} = -3.33$$

因为 $z < z_{\alpha = 0.05}$ 拒绝 H_0,接受 H_1,表明这一批产品零件的抗高温性能低于 1 250℃ 而不符合要求,因此不能接受这批产品。

(二) 未知总体分布及总体方差,大样本

在大样本的条件下,总体不论是否服从正态分布,由中心极限定理可知样本均值 $\bar{X}$ 近似地服从正态分布,其数学期望为总体均值,方差为总体方差 σ^2 的 $1/n$(n 为样本量)。总体方差未知时可用大样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 来估计,所以检验总体均值的统计量为:

$$Z = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \tag{6.2}$$

【例 6.2】 有一空调机的零件需用打孔机打孔,要求孔径为 10 厘米,太大太小都对装配有问题。为了测试打孔机是否正常,需要取样进行检验,在打孔的结果中随机取了 100 件进行测量,得 $\bar{x} = 9.6\text{cm}$, $s = 1\text{cm}$,试以 $\alpha = 0.05$,检验打孔机的操作是否正常,抑或如何调整。

解:先建立假设,由于检验结果过大或过小均不合适,因此应该是双侧检验,拒绝域在两侧。

$$H_0: \mu = 10 \text{ 厘米}, \quad H_1: \mu \neq 10 \text{ 厘米}$$

当 $\alpha = 0.05$ 时，临界值为 $\pm z_{\alpha/2} = \pm 1.96$。根据检验的结果 $z = \dfrac{9.6 - 10}{1/\sqrt{100}} = -4$，$z < -1.96$ 故拒绝 H_0 接受 H_1 说明孔径不等于 10 厘米而且是偏小，所以应对打孔机进行调整，使其孔径大一些。

假设检验与区间估计是相对应的，不难看出假设检验的接受域就是相应问题的置信区间，因此我们也可以用求置信区间的方法来作假设检验。就以例 6.2 的数据来说明，若 $\alpha = 0.05$，即要求以 95% 的置信水平来估计总体参数，已知 $\bar{x} = 9.6$ 在 H_0 为真的条件下 $\bar{X} \sim N\left(10, \dfrac{1}{100}\right)$，$\bar{X}$ 的标准差为 $\dfrac{1}{\sqrt{100}}$，置信系数为 ± 1.96，故 μ 的置信区间为

$$\mu \text{ 的下限} = \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 9.6 - 1.96 \frac{1}{\sqrt{100}} = 9.404$$

$$\mu \text{ 的上限} = \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 9.6 + 1.96 \frac{1}{\sqrt{100}} = 9.796$$

由于 μ 的上限仍然小于 10 厘米，上述置信区间不可能包括均值为 10 厘米，其结论也是拒绝 H_0，与前面假设检验的结论相同。

（三）总体为正态分布、方差未知、小样本的假设检验

当总体服从正态分布 $N(N, \sigma^2)$，μ 和 σ^2 都是未知参数，样本的容量又小于 30 时，要检验 H_0 时其检验的统计量为

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{6.3}$$

t 分布的自由度为 $n-1$，$S^2 = \dfrac{1}{n-1}\sum_{j=1}^{n}(X_i - \bar{X})^2$。

根据双侧检验还是单侧检验来确定拒绝域的临界值。当为双侧检验，显著性水平为 α 时，临界值为 $\pm t_{\alpha/2}$；当为右侧检验时，显著性水平为 α，临界值为 t_α；当为左侧检验时，显著性水平为 α，临界值为 $-t_\alpha$。

【例 6.3】 一个轮胎制造厂声称它的轮胎在正常行驶的条件下平均行驶里程至少在 40 000 公里以上,通常已知轮胎在正常行驶的条件下,其行驶里程数服从正态分布。某一推销商要随机抽取 15 个轮胎作试验,经过测试得到平均行驶里程为 42 000 公里,标准差为 5 000 公里,若显著性水平为 $\alpha=0.05$,能否从这些样本数据使该轮胎制造厂的声称得到证实。

解:这个问题中轮胎的行驶里程是愈多愈好,因此是一个单侧检验的问题。但是这个问题的假设可以有两种情况

假设 1:$H_0:\mu\geqslant 40\ 000$ 公里;$H_1:\mu<40\ 000$ 公里

假设 2:$H_0:\mu\leqslant 40\ 000$ 公里;$H_1:\mu>40\ 000$ 公里

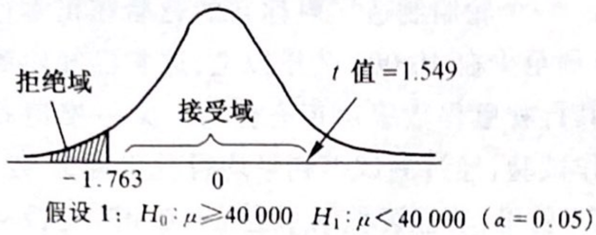
两种假设的结果将会得出两个截然不同的结论,我们不妨来计算一下,因为两种假设检验的统计量均是 t 统计量:

$$t=\frac{\bar{X}-\mu_0}{s/\sqrt{n}}=\frac{42\ 000-4\ 000}{5\ 000/\sqrt{15}}=1.549$$

假设 1 是左侧检验,临界值为 $-t_{\alpha=0.05,14}=-1.7613$ 拒绝域在临界值左侧,而 t 值落在右侧,因而是接受 H_0。假设 2 是右侧检验,临界值在右侧为 $t_{\alpha=0.05,14}=1.7613$,拒绝域在右侧而 t 值在临界值的左侧,也是接受 H_0,而这两个 H_0 的内容是不一样的,在假设 1 中是 $\mu\geqslant 40\ 000$ 公里,而在假设 2 中是 $\mu\leqslant 40\ 000$ 公里。究竟应该怎样来解释这种现象和选择哪一个假设,这是初学者在单侧假设检验时经常会遇到的问题。

这里我们可以先看一下两种假设情况下进行检验决策的图示,见图 6.5。

两种假设的目的是不同的,第一种假设中 $H_1:\mu<40\ 000$,是欲检验轮胎的平均行驶里程是否显著地低于 40 000 公里,而样本均值高于 40 000 公里,显然不用检验就可以接受 H_0,认为轮胎平均行驶里程没有显著地低于 40 000 公里。第二种假设中,$H_1:\mu>40\ 000$ 是欲检验轮胎的平均行驶里程是否显著高于 40 000 公



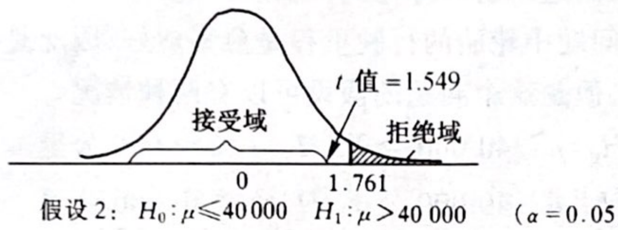


图 6.5 两种不同假设的决策图示

里，检验的结果接受 H_0，说明在 $\alpha=0.05$ 的显著性水平下，轮胎平均行驶里程也没有显著高于 40 000 公里。这样，两个假设就并不矛盾，说明既不显著低于 40 000 公里，又不显著高于 40 000 公里，它很可能接近于平均 40 000 公里。另外前面说过，接受 H_0 并不意味着 H_0 一定正确，而是说以现有显著性水平尚不能推翻 H_0，现在我们仍用假设 2 的形式，若把显著性水平改为 0.1，称作假设 3，这时拒绝域的临界值 $t_{\alpha=0.1}(14)=1.340\ 6$，因此 $t>t_\alpha$ 落入了拒绝域，结论变为接受 H_1，其图形如图 6.6 所示。

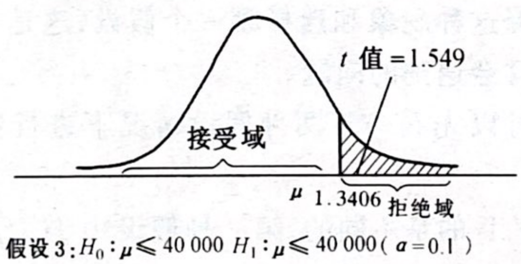


图 6.6 假设 3 图示

这样假设 2 与假设 3 由于显著性水平不同而得到了不同的结论。当显著性水平 $\alpha=0.05$ 时为接受 H_0，意味着尚不能以 95%

的置信水平推翻 H_0，当显著性水平 $\alpha=0.1$ 时接受 H_1，意味着有90%的置信水平推翻 H_0。

二、单一样本的比例检验

假如我们希望检验一个总体的比例 p 是否等于 p_0，显著性水平为 α。在检验时需要从总体中取一个容量为 n 的样本并计算样本比例 P。并假定 $np_0 \geqslant 5$，$nq_0 \geqslant 5(q_0=1-p_0)$。这些条件成立时，检验的统计量为

$$Z=\frac{P-p_0}{\sqrt{\frac{p_0 q_0}{n}}} \tag{6.4}$$

该统计量在原假设为真时近似地服从标准正态分布。若 $H_0: p=p_0$，$H_1: p \neq p_0$，决策规则为 $Z<-z_{\alpha/2}$，$Z>z_{\alpha/2}$ 时拒绝 H_0。若 $H_0: p \geqslant p_0$，$H_1: p<p_0$，决策规则为 $Z<-z_\alpha$ 时拒绝 H_0 接受 H_1。若 $H_0: p \leqslant p_0$，$H_1: p>p_0$，决策规则为 $Z>z_\alpha$ 时拒绝 H_0 接受 H_1

【例 6.4】 某西红柿酱生产厂向供应商购一批西红柿，规定若优质西红柿的比例在40%及以上按一般市场价格收购，若达不到此标准，应低于市场价格收购，现随机抽取了100个西红柿作检验，只有34个优质西红柿，样本比例 $P=34\%$ 因而欲按低于市场价格收购，但供应商认为样本比例不到40%，是随机原因引起的，试用显著性水平 $\alpha=0.05$ 进行检验并加以说明。

解：因为西红柿中优质的比例愈高愈好，主要是把不够标准的检验出来，因此是一个单侧检验问题。其次是根据研究的目的来建立原假设和备择假设，通常是想加以证实的问题放在备择假设 H_1，因为这时犯错误的概率 α 是可以知道的。这个例子中供应商相当于生产者，而西红柿酱生产厂相当于消费者，生产方总是怕将合格品当作不合格品而被拒收，因此要把产品显著低于标准才检验出来，把 $p<40\%$ 放在 H_1，即 $H_0: p \geqslant 40\%$，$H_1: p<40\%$，该例中 $n=100$ 为大样本，$np_0=40>5$，故应用 Z 统计量

$$Z=\frac{P-p_0}{\sqrt{\frac{p_0q_0}{n}}}=\frac{0.34-0.4}{\sqrt{\frac{(0.4)(0.6)}{100}}}=\frac{-0.06}{0.04899}=-1.23$$

$\alpha=0.05$，左侧检验临界值为 $-z_{\alpha=0.05}=-1.645$. 因此落入接受域，尚不能认为优质西红柿的比例显著地低于40%，仍应按市场价格收购。但是还须指出在接受 H_0 时并不意味着 H_0 一定是正确的，只是说明根据样本数据尚不能推翻 H_0。西红柿酱生产厂作为消费者在接受 H_0 时，潜在地存在第二类错误 β 的风险，而这一概率是未知的。如果要兼顾到生产者和消费者的双方利益，应在验收抽样设计中既规定 α 错误，又规定 β 错误。由于超出了本书要求，这里不作详细介绍。

三、两个总体均值之差的检验

有比较才有鉴别，客观事物总是在相互比较。如两种产品质量的比较，两个品种农作物产量的比较，两种农药杀虫效果的比较等等。两个总体均值之差也是研究和管理人员经常会遇到和感兴趣的参数。在对两个总体均值之差进行假设检验时，也需要通过两个样本的比较加以推断。其形式主要有以下三种：

$H_0:\mu_1-\mu_2=d_0, H_1:\mu_1-\mu_2\neq d_0$ （双侧检验）

$H_0:\mu_1-\mu_2\leqslant d_0, H_1:\mu_1-\mu_2> d_0$ （右侧检验）

$H_0:\mu_1-\mu_2\geqslant d_0, H_1:\mu_1-\mu_2< d_0$ （左侧检验）

在大多数情况下是假定 $d_0=0$，于是其形式是

$H_0:\mu_1=\mu_2 \qquad H_1:\mu_1\neq\mu_2$

$H_0:\mu_1\leqslant\mu_2 \qquad H_1:\mu_1>\mu_2$

$H_0:\mu_1\geqslant\mu_2 \qquad H_1:\mu_1<\mu_2$

至于检验的统计量，则与总体分布、方差是否已知和样本的容量大小有关。

1. 当两个正态总体方差已知和两个均为大样本时都采用 Z 检验：

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{6.5.a}$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{6.5.b}$$

【例 6.5】 一个新建的超市在选择位置时需考虑许多因素，因素之一是有关周围居民的收入水平，现有 A、B 两地可供选择，A 地的建筑费用较 B 地低，如果两地居民的平均收入相同，就在 A 地建筑，但若 B 地的居民平均收入高于 A 地则选在 B 地建筑。现从两地的居民户中各抽取了 100 户居民，调查并计算其收入水平。A 地年平均收入 28 650 元，从其他方面获知总体标准差 4 740 元，B 地年平均收入 29 980 元，获知标准差为 5 365 元。用 $\alpha=0.05$ 的显著性水平推断 B 地的收入水平是否显著高于 A 地，然后决策在何地建筑超市。

解：目的是要通过样本比较两个地区居民总体的平均收入之差，要检验的参数是 $\mu_B-\mu_A$，μ_B 表示 B 地区的年平均收入，μ_A 表示 A 地区年平均收入。

$$H_0: \mu_B - \mu_A \leqslant 0, \quad H_1: \mu_B - \mu_A > 0$$

由于两个样本均为大样本且方差已知，可用检验统计量公式(6.5.a)，$\alpha=0.05$，右侧检验，决策规则为 $Z > z_{\alpha=0.05}=1.645$ 时拒绝 H_0 接受 H_1，

$$Z = \frac{(\bar{x}_B - \bar{x}_A) - (\mu_B - \mu_A)}{\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}}} = \frac{29\,980 - 28\,650}{\sqrt{\frac{5\,365^2}{100} + \frac{4\,740^2}{100}}} = 1.86$$

$Z>1.645$ 拒绝 H_0 接受 H_1，表明 B 地的居民平均年收入高于 A 地，应在 B 地建设超市。

【例 6.6】 某大学欲比较大学毕业后留在学校工作与分配到其他工作岗位的人工资水平的差别，因为工资还与工龄等其他因

素有关，因此抽选大学毕业后满 10 年在校工作的教师 50 人，另外抽选大学毕业后满 10 年在机关、企业工作的人员 50 人进行比较，取得的数据如下：

大学教师	机关、企业工作人员
$n_1 = 50$	$n_2 = 50$
$\bar{x}_1 = 23\ 700$	$\bar{x}_2 = 21\ 500$
$s_1 = 2\ 435$	$s_2 = 6\ 804$

试比较大学毕业后留校当教师与分配在机关、企业等工作人员的工资水平是否有差别（$\alpha = 0.05$）。

解：比较工资水平的高低，可用双侧检验，由于两个总体的样本 $n_1 = n_2 = 50$ 为大样本，但总体方差未知，因此可用检验统计量公式(6.5.b)

$$H_0: \mu_1 = \mu_2, \qquad H_1: \mu_1 \neq \mu_2 \quad (D_0 = 0)$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(23\ 700 - 21\ 500) - 0}{\sqrt{\dfrac{2\ 435^2}{50} + \dfrac{6\ 804^2}{50}}} = \frac{2\ 200}{1\ 022} = 2.15$$

当 $\alpha = 0.05$ 时，由于是双侧检验，拒绝域在两侧，临界值为 ± 1.96，$Z > 1.96$，拒绝 H_0 接受 H_1，说明收入不相等。进一步从样本数据分析是大学留校的教师平均工资高于其他岗位，但从标准差看，其他岗位的差别大于大学教师（但未作检验）。

2. 比较两个正态总体、方差未知小样本，则需 t 检验统计量：

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{S_{合}^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \tag{6.6}$$

该统计量假定两个总体的方差相等，用两个样本数据合并计算，其中：

$$S_{合}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

t 检验统计量的自由度为 $n_1 + n_2 - 2$。

【例 6.7】 工厂管理人员对组装新产品的两种方法所需的时间进行测试,他们认为顺序的合理与否是一个关键,顺序合理就能节省时间提高效率。随机抽选采用方法 A 的 6 个工人,和采用方法 B 的 8 个工人,测试的结果如下:

方法 A	8.2	5.3	6.5	5.1	9.7	10.8		
方法 B	9.5	8.3	7.5	10.9	11.3	9.3	8.8	8.0

假设组装的时间服从正态分布,以 $\alpha = 0.05$ 显著性水平比较两种组装方法有否差别。

解:本例是两个正态总体、方差未知、小样本情形,适合用 t 检验统计量的公式(6.6)。

设方法 A 所需的平均时间为 μ_1,方法 B 所需的平均时间为 μ_2。

$$H_0: \mu_1 = \mu_2, \quad H_1: \mu_1 \neq \mu_2 \quad (d_0 = 0)$$

根据样本获得的数据计算得

$$\bar{x}_1 = 7.6 \quad s_1 = 2.36, \bar{x}_2 = 9.2, s_2 = 1.35,$$

计算合并方差

$$s_{合}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{5(2.36)^2 + 7(1.35)^2}{6 + 8 - 2} = 3.38$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_{合}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(7.6 - 9.2) - 0}{\sqrt{3.38\left(\frac{1}{6} + \frac{1}{8}\right)}} = -1.61$$

当 $\alpha = 0.05$ 临界值为 $\pm t_{\alpha/2, n_1 + n_2 - 2} = \pm 2.179$,$t$ 落入接受域,故接受 H_0,表明方法 A 与方法 B 无显著差别。

四、两个总体比例之差的检验

两个总体比例之差的检验与两个总体均值之差的检验一样，只是比较的两个总体都是(0－1)分布的总体，即两个总体中具有某种特征单元数的比例进行比较。另外由于 p 的方差计算有其特点，使检验统计量的公式也略有变化。现分两种情况加以讨论。

1. 检验两个总体比例之差是否等于 0，即 $H_0: p_1 - p_2 = 0$，$H_1: p_1 - p_2 \neq 0$。很自然地样本的估计量为两个样本比例之差，因此其抽样分布的数学期望和方差为：

$$E(P_1 - P_2) = E(P_1) - E(P_2) = p_1 - p_2$$

$$D(P_1 - P_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

在原假设为真时(即 $p_1 = p_2$)，其方差应为 $\frac{pq}{n_1} + \frac{pq}{n_2} = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$，但 pq 为待估参数也需要用样本来估计。

若 $P_1 = \frac{X_1}{n_1}, P_2 = \frac{X_2}{n_2}$，则合并的估计值 $\hat{P} = \frac{X_1 + X_2}{n_1 + n_2}$(即总共 $n_1 + n_2$ 个样品中有 $X_1 + X_2$ 个样品具有指定的特征)，因此检验的统计量为

$$Z = \frac{(P_1 - P_2) - (p_1 - p_2)}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (6.7)$$

以上检验要求两个比例均为大样本，且有 $n_1 P_1, n_1 Q_1, n_2 P_2, n_2 Q_2$ 都大于 5。

【例 6.8】 某保险公司在人寿保险中拟对不抽烟的人在交纳保险时提供折扣，作了一项抽烟人犯心脏病的比例和不抽烟的人犯心脏病的比例的比较，调查的对象是 50 岁的男性，抽烟的人是指每天至少抽一包的人，分别问他们是否犯过心脏病。调查结果 80 名抽烟者中有 20 名得过心脏病，120 名不抽烟的人中有 15 人得过心脏病。要求以 5% 的显著性水平推断抽烟人与不抽烟的人

犯心脏病的比例是否有显著差别。

解：令 p_1——抽烟人中犯过心脏病的比例；

p_2——不抽烟的人中犯过心脏病的比例；

n_1——抽烟人的样本量；

n_2——不抽烟人的样本量；

x_1——抽烟样本中犯过心脏病人数；

x_2——不抽烟样本中犯过心脏病人数；

P_1——x_1/n_1，抽烟样本中犯过心脏病的比例；

P_2——x_2/n_2，不抽烟样本中犯过心脏病的比例。

公司研究的目的是要了解抽烟群体中犯过心脏病人的比例是否显著高于不抽烟的群体，因此是一个单侧的假设检验。

$$H_0: p_1 - p_2 \leqslant 0, \qquad H_1: p_1 - p_2 > 0$$

将要检验的目的放在 H_1，这样在拒绝 H_0 接受 H_1 时可能犯错误的概率 α 是可以预先规定的。在本例中 $\alpha = 0.05$。

根据提供的数据

$$P_1 = \frac{x_1}{n_1} = \frac{20}{80} = 0.25,$$

$$P_2 = \frac{15}{120} = 0.125,$$

$$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{35}{200} = 0.175$$

检验的统计量

$$Z = \frac{(P_1 - P_2) - (p_1 - p_2)}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(0.25 - 0.125) - 0}{\sqrt{(0.175)(0.825)\left(\frac{1}{80} + \frac{1}{120}\right)}} = 2.28$$

在 $\alpha = 0.05$ 时，拒绝域在右侧，临界值 $z_\alpha = 1.645$，$Z > z_\alpha$，拒绝 H_0 接受 H_1，表明抽烟的人中犯过心脏病的比例要超过不抽烟的

人。为公司要对不抽烟的人提供折扣提供了依据。

2．检验两个总体比例之差不等于0，即

$$H_0: p_1 - p_2 = d_0, \quad H_1: p_1 - p_2 \neq d_0$$

这时即使原假设 H_0 为真时，$p_1 \neq p_2$，因此对于总体方差不必合并计算，因此检验的统计量为：

$$Z = \frac{(P_1 - P_2) - d_0}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}} \tag{6.8}$$

【例6.9】 有两种方法生产同一种电冰箱零件，分别称为方法一、方法二。方法一的生产费用较高而次品率比较低，方法二生产费用低而次品率高。厂方管理人员在衡量两种方法的利弊时决定对两种方法的次品率作一比较，若第一种方法比第二种方法的次品率低8%以上，可以采用方法一，否则就采用方法二。于是用两种方法进行试验，然后从方法一的产品中随机抽取了300个零件，发现有33个次品，从方法二的产品中也随机抽取了300个零件，发现有84个次品。试以1%的显著性水平进行假设检验，说明决定用哪一种方法进行生产。

解：设 p_1 为第一种方法次品率，p_2 为第二种方法次品率。

$$H_0: p_2 - p_1 \leqslant 0.08, H_1: p_2 - p_1 > 0.08$$

由于 p_1 和 p_2 不相等，对总体比例不必合并计算，又 $n_1 = n_2 = 300$ 均为大样本，因此适用 Z 检验统计量公式(6.8)，由 $P_1 = \dfrac{33}{300} = 0.11, P_2 = \dfrac{84}{300} = 0.28$

$$Z = \frac{(P_1 - P_2) - d_0}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}} = \frac{(0.28 - 0.11) - 0.08}{\sqrt{\dfrac{(0.11)(0.89)}{300} + \dfrac{(0.28)(0.72)}{300}}}$$

$$= \frac{0.09}{0.0315} = 2.85$$

这是单侧检验。当 $\alpha = 0.01$ 时，$z_{\alpha=0.01} = 2.33$，$Z > z_\alpha$，故拒绝 H_0，接受 H_1，表明方法二的次品率显著高于方法一8%，故可采用方法一。