第 7 章

参数假设检验

假设检验是用来判断样本与样本、样本与总体的差异是由抽样误差引起还是本质差别造成的统计推断方法。其基本原理是先对总体的特征作出某种假设，然后通过抽样研究的统计推理，对此假设应该被拒绝还是接受作出推断。假设检验可分为参数假设检验和非参数假设检验。本章主要介绍参数假设检验的思想与步骤、正态总体单样本参数假设检验、正态总体双样本参数假设检验和比例假设检验。

7.1 假设检验的思想与步骤

在本节，我们首先介绍假设检验的思想，然后再介绍假设检验的基本步骤。

7.1.1 假设检验的基本思想

关于假设检验我们先看一个经典的女士品茶问题。

例7.1　在20世纪20年代后期，英国剑桥一个夏日的午后，一群大学的绅士和他们的夫人们，享用着下午茶。在品茶过程中，一位女士坚称：把茶加进奶里，或把奶加进茶里，不同的做法，会使茶的味道品起来不同。在场的一帮科学精英们，对这位女士的“胡言乱语”嗤之以鼻。这怎么可能呢？然而，在座的一个身材矮小、戴着厚眼镜、下巴上蓄着短胡须的费希尔先生，却不这么看，他对这个问题很感兴趣。他兴奋地说道：“让我们来检验这个命题吧！”并开始策划一个实验。在实验中，坚持茶有不同味道的那位女士被奉上10杯已经调制好的茶，其中，有的是先加茶后加奶制成的，有的则是先加奶后加茶制成的。接下来，这位先生不加评论地记下了女士的说法，结果此女士准确地分辨出了10杯中的每一杯。

给出两种假设，分别是原假设H_0和与之相反的备择假设H_1。在这个问题中我们做如下假设：

原假设H_0：该女士没有此种鉴别能力。

备择假设H_1：该女士有此种鉴别能力。

在原假设成立的前提下，即每一杯茶她有0.5的概率猜对。10杯茶全部猜对的概率是$(1/2)^{10}$，这是一个非常小的概率，认为在一次实验中不会发生，但这件事情发生了，只能说明原假设不当，应该拒绝，从而认为该女士有鉴别能力。

若10杯中只有6杯说对了，该如何判断这个问题呢？

若100杯中有60杯说对了，又该如何判断这个问题呢？

女士品茶鉴别情况				
总　量	判断正确	判断错误	假设正确概率	是否拒绝假设
10	10	0	0.5	拒绝
10	6	4	0.5	?
100	60	40	?	?

假设检验是常用的一种统计推断方法。其基本思想是小概率反证法思想。将待检验的问题分为两个相互矛盾的假设，分别为原假设H_0和备择假设H_1。先假定原假设H_0成立，并在原假设成立的条件下，建立相应的枢轴量。通过枢轴量的分布，划定小概率区间。若已发生样本对应的枢轴量落于小概率区间内，则认为有理由拒绝原假设，而接受备择假设，因为小概率事件往往不发生。反之，则不能拒绝原假设，但往往不能说就接受原假设。因为或许是数据量小了，导致小概率事件没有发生，这时没有理由确信原假设一定正确。但当样本量很大时，则不能拒绝原假设，往往我们说“可以接受原假设”。

假设检验分为参数假设检验和非参数假设检验。本章首先介绍的参数假设检验，即总体分布类型已知，用样本指标对总体参数进行推断的统计分析方法。关于非参数统计，不对总体的分布类型进行假设，即所判断的假设不涉及总体参数的统计推断。

在数学推导上，参数假设检验是与区间估计是相联系的，而在方法上，二者又有区别。对于区间估计，人们主要是通过数据推断未知参数的取值范围；而对于假设检验，人们则是做出一个关于未知参数的假设，然后根据观察到的样本判别该假设是否正确。在R中，区间估计和假设检验使用的是同一个函数。

7.1.2　假设检验的基本步骤

我们先看一个例子，来说明假设检验的目的及步骤。

例7.2　某公司生产轮轴，直径均值为5.00cm，标准差是1.00cm，假定轮轴的直径服从正态分布，标准差是1.00cm。目前你作为该公司的检验员，通过抽取的样本（$X_1, X_2, \ldots, X_n$），如表7-1所示，如何判定均值就是5.00cm呢？

表7-1　轮轴样本数据

4.89	5.99	5.89	6.22	4.79	5.47	4.50	6.61	4.25	6.67
4.46	4.50	6.97	5.39	4.56	5.03	2.54	5.27	4.48	4.05

这是假设检验问题，即判别某一假设是否正确。

首先给出两个相互矛盾的假定。

原假设H_0：轮轴直径达到5.00cm　$\mu = \mu_0 = 5$

备择假设H_1：$\mu \neq \mu_0 = 5$

其次构造枢轴量：$u = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$，根据中心极限定理，可以证明在$H_0$成立时，$\mu$~$N(0,1)$。所以$\mu$落在图中阴影覆盖部分时是小概率事件，概率为$\alpha$（$\alpha$被称作显著性水平，是指小概率区间的概率大小，一般取值为0.05或者0.01）。比如取α=0.05。若$-1.96<\mu<1.96$，则说明落在接受域，不能拒绝H_0，若$\mu<-1.96$或者$\mu>1.96$，则拒绝原假设H_0，接受H_1，如图7-1所示。

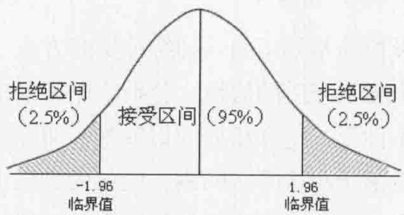


图7-1　假设检验示意图

关于该检验，我们可以编写一个做双边检验的函数u.test()进行检验。

```
> u.test<-function(a,mu,thegma)
{   Se=thegma/sqrt(length(a))
    u=(mean(a)-mu)/Se
    p=2*(1-pnorm(abs(u)))
    return(list(u=u, p=p))
}
> a<-c(4.89,5.99,5.89,6.22,4.79,5.47,4.50,6.61,4.25,6.67,4.46,4.50,6.97,5.39,
  4.56,5.03,2.54,5.27,4.48,4.05)
>u.test(a,5,1)
$u
[1] 0.5657252
$p
[1] 0.5715806
```

检验的u统计量是0.5657252，对应的p值是0.5715806，即在H_0成立的前提下，u统计量为0.5657252的概率为0.5715806。说明不能拒绝原假设，也就是没有证据认为轮轴直径不是5.00cm。

综合以上，我们可以看出假设检验分为以下几个步骤。

① 建立两个互斥的假设分别设为原假设 H_0 和备择假设 H_1。注意，若这两个假设互换，则可能会导致结论有差异，因为假设检验原理其实一个反证法，只能明确地说明原假设不成立，但往往不能证明原假设是对的。

② 找到合适的枢轴量。上面例子中的 u 在原假设成立的条件下，分布已知。这时候可以依据其分布给出小概率区间，即拒绝域。

③ 通过样本计算枢轴量，做出判别，或者给出p值。对于给定显著性水平 α 时，通过样本计算出来的 u 值，当落入拒绝域时，接受备择假设，否则不能拒绝原假设。

但显著性水平 α 如何选取？在原假设成立的条件下，拒绝原假设称之为第一类错误。当原假设不正确时，却没能拒绝原假设称之为第二类错误。α 越大，第一类错误越大，α 越小，第二类错误越大。要同时减小两类错误，则只能增加样本量，在现实生活中增加样本量有时候是可行的，有时候是不可行的。例如，上面的女士品茶实验，当样本量增加以后，女士的口感也就弱化了，本来有可能有鉴别能力，但大量增加品茶次数之后，鉴别能力或许就没有了。

根据刚才的分析，显著性水平 α 的大小要根据具体研究问题来选取。通常取为0.05。在很多统计软件中，为了避免显著性水平 α 的大小的选取，取而代之的是给出 μ 值相应的p值，让读者自己取舍。

7.2　正态总体单样本参数假设检验

自然界和人类社会中很多变量服从或者近似服从正态分布，本节以随机变量服从正态分布

为前提，介绍对正态总体参数的假设检验。本节先介绍单个正态总体的参数假设检验问题。

7.2.1 均值的检验

1. 方差已知情形

例7.3 某汽车生产商声称其生产的汽车每加仑汽油可行驶的里程不低于25英里，标准差为2.4英里。消协组织了一个由10位汽车主组成的小组，他们的汽车每加仑汽油的可行驶英里数如下表。假定汽车每加仑可行驶里程服从正态分布。则汽车生产商的诺言可信吗？

序号	1	2	3	4	5	6	7	8	9	10
里程	22	24	21	24	23	24	23	22	21	25

方差已知情形的理论推导和检验步骤如下。

设 $x_1,x_2,...,x_n$ 来自 $X\sim N(\mu,\sigma^2)$，σ^2 已知，对均值 μ 进行检验。

（1）检验假设。H_0：$\mu=\mu_0$ 和 H_1：$\mu\neq\mu_0$。

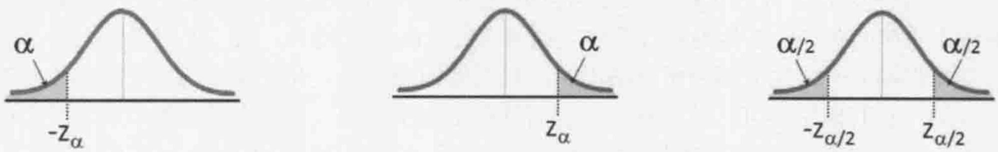
（2）给定检验水平 α。

（3）枢轴量 $u=\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$ 服从标准正态分布。

（4）计算 u 统计量对应的p值。

（5）若 u 值落在拒绝域内，则拒绝 H_0，反之不能拒绝 H_0。

容易推导，若原假设变化时，拒绝域的变化如图7–2所示。



（a） H_1：$\mu<\mu_0$

（b） H_1：$\mu>\mu_0$

（c） H_1：$\mu\neq\mu_0$

图7–2 不同原假设下的拒绝域

通常，我们将拒绝域所在区域与之检验相对应。第三种情况称之为双边检验，前面两种都是单边检验，分别又称为左侧检验和右侧检验。

回到刚才的问题，需要检验的两个假设是 H_0：$\mu\geq\mu_0$ 和 H_1：$\mu<\mu_0$。

可以构造枢轴量：$u=\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$。在 $\mu=\mu_0$ 时， 服从 $N(0,1)$。也就是说，在原假设成立的条件下，u 落在图7-3拒绝区域中的概率会比较小。通过计算 u 值对应的 p 值。若 $p\le\alpha$，则拒绝H_0，接受H_1。若 $p>\alpha$，则接受H_0，拒绝H_1。

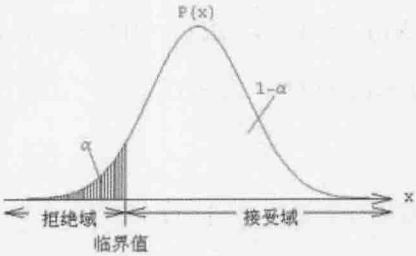


图7-3 左侧检验示意图

此例中H_0: $\mu\ge 25$，备择假设H_1为 $\mu<25$。前面我们自己编写过u.test()函数进行双边检验，此处可以在前文编写的u.test()函数基础上进行修改，使之不仅能做双边检验，也能做左侧单边检验和右侧单边检验。

```
u.test<-function(a,mu,thegma,alternative="twoside")
{  Se=thegma/sqrt(length(a))
   u=(mean(a)-mu)/Se
   if (alternative=="twoside") p=2*(1-pnorm(abs(u)))
   else if (alternative=="less") p=pnorm(u)
   else   p=1-pnorm(u)
   return(list(u=u, p=p))
}
> b=c(22,24,21,24,23,24,23,22,21,25)
>u.test(b,25,2.4,alternative="less")  #左侧检验
$u
[1] -2.766993
$p
[1] 0.002828799
```

通过p值可以看出，落在拒绝域内，可以认为厂家承诺没有达到。

2. 方差未知情形

例7.4 一位投资者正在考虑是否选择新的资产管理公司，为了使收益最大化，如果新的资产管理公司平均收益率大于原来资产管理公司的平均收益率，则公司将选择新的资产管理公司。原来的资产管理公司的客户平均收益率为50.0%，对新资产管理公司的客户进行抽样检验，12个客户的收益率如下：50.2%、49.6%、51.0%、50.8%、50.6%、49.8%、51.2%、49.7%、51.5%、50.3%、51.0%和50.6%，假设资产管理公司客户收益率的分布比较近似于正态

分布，则新资产管理公司的平均收益率是否大于原来的资产管理公司?

这个问题和上面已处理的问题很相似，只是不知道方差取值多少。在操作时步骤也是一样的，只是选取的枢轴量不同。

方差未知时的理论推导和检验步骤如下。

设$x_1, x_2, ..., x_n$来自$X \sim N(\mu, \sigma^2)$，σ^2未知，对均值μ进行检验。

（1）检验假设。H_0：$\mu = \mu_0$和H_1：$\mu \neq \mu_0$。

（2）给定检验水平α。

（3）枢轴量$t = \dfrac{\overline{x} - \mu_0}{s / \sqrt{n}}$服从自由度为$n-1$的$t$分布$t(n-1)$。

（4）计算 值对应的p值。

（5）若t值落在拒绝域内，则接受H_1，反之不能拒绝H_0。

容易推导，若原假设变化时，拒绝域的变化同上，当原假设是$\mu = \mu_0$时是双边检验，其他两种情况是单边检验。

本例中设H_0：$\mu \le \mu_0$和H_1：$\mu > \mu_0$，取显著性水平α=0.05。

下面用R的内部函数t.test()进行检验：

```
> x=c(50.2,49.6,51.0,50.8,50.6,49.8,51.2,49.7,51.5,50.3,51.0,50.6)
>t.test(x,mu=50,alternative ="greater") #H0: μ0≤μ1和H0: μ0>μ1
     One Sample t-test
data:  x
t = 2.9564, df = 11, p-value = 0.006529
alternative hypothesis: true mean is greater than 50
95 percent confidence interval:
 50.20609      Inf
sample estimates:
mean of x
   50.525
```

检验的p = 0.0065≤0.05，在显著性水平α=0.05时拒绝H_0，接受H_1，可认为新资产管理公司的收益率的确更高。

7.2.2 方差检验

方差的大小表现的是总体的离散程度，在现实生活中，很多时候需要对方差进行检验。

例7.5 某地区环保部门规定，废水经设备处理后，水中某种有毒物质的平均浓度为

$X \sim N(500, 20^2)$，均值和标准差单位是微克。通过抽查20个废水样品，如何判断废水处理设备是否正常工作？数据见R程序中变量x。

对于这个问题，需要检验均值是否为500ug和标准差是否为20ug。若这两个指标有差异，则不能说明机器正常生产，会导致废水中有害物质浓度过高或不能均匀分布从而影响有害物质的降解。关于均值的检验在7.2.1节已经介绍过了，本节将讨论标准差的检验。

理论推导和检验步骤如下。

设 $x_1, x_2, ..., x_n$ 来自 $X \sim N(\mu, \sigma^2)$，对方差 σ^2 是否与 σ_0^2 相等进行检验。

（1）检验假设。$H_0: \sigma^2 = \sigma_0{}^2$ 和 H_1： $\sigma^2 \neq \sigma_0{}^2$。

（2）给定检验水平 α。

（3）枢轴量 $\chi^2 = (n-1)s^2 / \sigma_0{}^2$ 服从 $\chi^2(n-1)$。

（4）计算 χ^2 值对应的 p 值。

（5）若 χ^2 值落在拒绝域内，则接受 H_1，反之不能拒绝 H_0。

与前面推导相似，若原假设变化时，拒绝域的变化同上；当原假设是 $\sigma^2 = \sigma_0^2$ 时是双边检验，其他两种情况是单边检验。

R实战演习

```
> x=c(512.952899108198, 503.85274864927, 495.06951127009, 477.193305294993,
509.400520346022, 493.249014260413, 492.456674317536, 530.078195416527,
498.757258963417, 522.657000900506, 510.041124496973, 490.505063978937,
536.24855605503, 530.039965979141, 495.559165160148, 466.840695851664,
510.702680801237, 524.485012890925, 490.37616974915, 485.579333872921
)
> var.test1<-function(x, sigma2){
  n<-length(x)
  S2=var(x)
df=n-1
  chi2<-df*S2/sigma2;
  P<-pchisq(chi2,df)
data.frame(var=S2, df=df, chisq2=chi2, P_value=P)
}
> var.test1(x,400)
vardf   chisq2       P_value
    346.8209   19    16.47399      0.3745438
```

通过 p 值可以知道 χ^2 没有落在拒绝域内，不能说明机器工作不正常，也就是默认机器正常工作。

7.3 正态总体双样本参数假设检验

现实生活中，往往需要比较两个正态总体的样本参数是否相同。由于均值的检验中需要用到方差的检验，所以本节先介绍方差的检验，然后再介绍均值检验。

7.3.1 双样本方差的检验（方差齐性检验）

例7.6　假设有一个轮胎生产商，需从两个厂家买入轮轴，假设每个轮轴的直径均服从正态分布，设总体 $X_1 \sim N(\mu_1,\sigma_1^{\ 2})$ 和 $X_2 \sim N(\mu_2,\sigma_2^{\ 2})$，请问对轮轴的方差是否有要求，是越小越好，还是越大越好？如何检验两者方差大小？具体样本数据见R程序中X_1、X_2变量。

理论推导和检验步骤（F检验）

用数学语言描述，总体 $X_1 \sim N(\mu_1,\sigma_1^{\ 2})$ 和 $X_2 \sim N(\mu_2,\sigma_2^{\ 2})$，$X_1$ 与 X_2 相互独立，记 $s_1^{\ 2}$ 和 $s_2^{\ 2}$ 分别为两个样本的方差，n_1 和 n_2 分别为样本容量，问题是如何检验两总体方差 $\sigma_1^{\ 2}$ 与 $\sigma_2^{\ 2}$ 相等，或者说哪个明显更大一些？

对 $\sigma_1^{\ 2}$ 与 $\sigma_2^{\ 2}$ 是否相等进行检验。

（1）检验假设。H_0：$\sigma_1^{\ 2}=\sigma_2^{\ 2}$ 和 H_1:$\sigma_1^{\ 2} \neq \sigma_2^{\ 2}$。

（2）给出显著性水平 α。

（3）枢轴量：$F=\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F(n_1-1,n_2-1)$。

　　在 H_0 成立时，统计量 $F=\dfrac{s_1^2}{s_2^2} \sim F(n_1-1,n_2-1)$。

（4）计算 F 值对应的p值。

（5）若 F 值落在拒绝域内，则接受 H_1，反之不能拒绝 H_0。

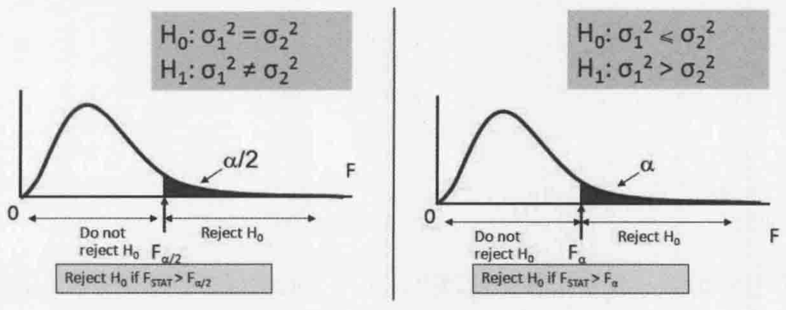


图7-4　不同原假设下的检验拒绝域

同前面推导相似，若原假设变化时，拒绝域的变化同上，当原假设 $\sigma_1^2=\sigma_2^2$ 时是双边检验，其他两种情况是单边检验。

R实战演习

```
> x1=c(24, 29, 39, 40, 32, 32, 31, 44, 37, 37, 50, 28, 24, 48, 25,
40, 32, 34, 35, 41)
> x2=c(44, 34, 36, 38, 30, 30, 35, 38, 40, 46, 38, 35, 38, 36, 38,
40, 34, 37, 40, 46)
> var.test(x1,x2)
     F test to compare two variances

data:  x1 and x2
F = 2.9283, numdf = 19, denomdf = 19, p-value = 0.02385
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.159058 7.398216
sample estimates:
ratio of variances
          2.928304
```

在显著性水平0.05条件下，F检验统计量为2.9283，对应的检验p值为0.02385，拒绝原假设，因此可以认为两者方差有明显的差异。

7.3.2　两样本均值检验

两样本检验为将一个样本与另一样本均值相比较的检验，在分析上和单样本检验类似，但在计算上有一些区别。两样本均值检验分为两独立样本检验和配对样本检验。两者适用条件不同，两独立样本检验适用于两个样本来源是相互独立的，配对样本检验则适用于两个样本是配对样本。

1. 两独立样本t检验

例7.7　一员工对乘当地公交车上班快还是自己开车快的问题产生了兴趣。通过对两种方式所用时间各进行了10次记录，具体数据见下表。设每一种方式的天数是随机选取的，假设乘车时间服从正态分布。试按下列要求进行分析，这些数据能提供充分的证据说明开车去上班的平均时间更快吗？用显著水平5%，并考虑用单尾检验还是双尾检验。

公交	48	47	44	45	46	47	43	47	42	48
开车	36	45	47	38	39	42	36	42	46	35

理论推导和检验步骤

要具体检验以下假设。

$$H_0: \mu_1 = \mu_2 \text{ 和 } H_1: \mu_1 \neq \mu_2$$

① 方差齐性时。

方差齐性的情况下理论推导和检验步骤。

两个总体的方差相等，$\sigma_1=\sigma_2=\sigma$，设 $x_1,x_2,...,x_n$ 来自总体 $X \sim N(\mu_1,\sigma^2)$，$y_1,y_2,...,y_m$ 来自总体 $Y \sim N(\mu_2,\sigma^2)$，其中 m 和 n 为两个样本的样本容量。

对 μ_1 是否与 μ_2 相等进行检验。

（1）检验假设。H_0：$\mu_1 = \mu_2$；H_1：$\mu_1 \neq \mu_2$

（2）给出显著性水平 α。

（3）枢轴量：$t=\dfrac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{s_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$，其中 $s_w=\dfrac{1}{m+n-2}[\sum_{i=1}^{m}(x_i-\overline{x})+\sum_{i=1}^{n}(y_i-\overline{y})]$，

在 H_0 成立时，统计量 $t=\dfrac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{s_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}} \sim t(m+n-2)$。

（4）计算 t 值对应的 p 值。

（5）若 t 值落在拒绝域内，则接受 H_1，反之不能拒绝 H_0。

当原假设变化时，拒绝域也相应的变化：当原假设是 $\mu_1=\mu_2$ 时是双边检验，其他两种情况是单边检验。

② 方差不齐时。

方差不齐时的理论推导和检验步骤。

方差不齐时，设 $x_1,x_2,...,x_n$ 来自总体 $X \sim N(\mu_1,\sigma_1^2)$，$y_1,y_2,...,y_m$ 来自总体 $Y \sim N(\mu_2,\sigma_2^2)$。

对 μ_1 与 μ_2 是否相等进行检验。

（1）检验假设。H_0：$\mu_1=\mu_2$ 和 H_1：$\mu_1 \neq \mu_2$

（2）给出显著性水平 α。

（3）枢轴量：$t=\dfrac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{s_x^2}{m}+\dfrac{s_y^2}{n}}}$，$s_x^2$ 为来自总体X样本的样本方差，s_y^2 为来

自总体Y样本的样本方差，在 H_0 成立时，统计量 $t=\dfrac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{s_x^2}{m}+\dfrac{s_y^2}{n}}}\sim t(l)$ ，其中

$l=(\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n})^2/(\dfrac{s_1^4}{m^2(m-1)}+\dfrac{s_2^4}{n^2(n-1)})$。

（4）计算 t 值对应的 p 值。

（5）若 t 值落在拒绝域内，则接受 H_1，反之不能拒绝 H_0。

当原假设变化时，拒绝域也相应的变化：当原假设是 $\mu_1=\mu_2$ 时是双边检验，其他两种情况是单边检验。

③ R实战演习。

（1）在做两样本均值检验时，需要验证样本是否服从正态分布，即正态性检验。（在后面内容介绍。）

（2）判断两个样本是否有相同的方差，可以根据方差齐次检验判别（前面已经介绍）。

```
> x1=c(48,47,44,45,46,47,43,47,42,48)
> x2=c(36,45,47,38,39,42,36,42,46,35)
> var.test(x1,x2)

	F test to compare two variances

data:  x1 and x2
F = 0.2273, numdf = 9, denomdf = 9, p-value = 0.03793
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.05646413 0.91520616
sample estimates:
ratio of variances
         0.2273243
```

$p=0.0379<0.05$，说明两组数据的方差是不一样的。

（3）t检验判断均值。

对于R软件中，两个独立样本检验的两种情况都用一个函数，即t.test()。用法如下：

```
> t.test (x1,x2,var.equal=T)   #方差齐次条件满足时
> t.test (x1,x2)               #默认方差非齐次时
```

默认方差非齐性，如果要假定方差齐性，则使用t.test()时要设定var.equal=TRUE。

```
> t.test (x1,x2)
	Welch Two Sample t-test
```

```
data:  x1 and x2
t = 3.288, df = 12.89, p-value = 0.005939
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.746 8.454
sample estimates:
mean of x mean of y
      45.7      40.6
```

经检验，$p= 0.0059 < 0.05$，拒绝原假设，说明自己开车和坐公交车所花的时间不同，自己开车所花时间较少。

2. 两配对样本t检验

配对或成对样本t检验使用的是不一样的统计量。配对样本检验假定两样本有相同的一些属性，而不是假定它们是独立正态分布的。

基本的模型是$Y_i = X_i + \varepsilon_i$，其中$\varepsilon_i$为随机项。我们想检验$\varepsilon_i$的均值是不是0，为此用$Y$减去$X$，然后做通常的单样本$t$检验。

例7.8　一个以减肥为主要目标的健美俱乐部声称，参加其训练班至少可以使肥胖者平均体重减轻8.5kg以上。为了验证该宣传是否可信，调查人员随机抽取了10名参加者，得到他们的体重记录如下。

训练前	94.5	101	110	103.5	97	88.5	96.5	101	104	116.5
训练后	85	89.5	101.5	96	86	80.5	87	93.5	93	102

理论推导和检验步骤

要具体检验以下假设。

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

在正态性假定下，$d = X - Y$近似服从$N(\mu, \sigma_d^{\ 2})$。其中，$\mu = \mu_1 - \mu_2$，$\sigma_d^{\ 2} = \sigma_1^{\ 2} + \sigma_2^{\ 2}$。需要比较$\mu_1$与$\mu_2$大小的问题转变成了$\mu$是否为0。

对均值μ是否为0进行检验。

（1）检验假设。H_0：$\mu = 0$ 和 H_1：$\mu \neq 0$

（2）给出显著性水平α。

（3）枢轴量：$t = \overline{d} / (s_d / \sqrt{n})$，其中$d_i = x_i - y_i$，$\overline{d} = \dfrac{1}{n}\sum_{i=1}^{n} d_i$，

$$s_d=[\frac{1}{n-1}\sum_{i=1}^{n}(d_i-\overline{d})^2]^{\frac{1}{2}}$$

在H_0成立时，统计量$t=\overline{d}/(s_d/\sqrt{n})\sim t(n-1)$。

（4）计算t值对应的p值。

（5）若t值落在拒绝域内，则接受H_1，反之不能拒绝H_0。

若原假设变化时，拒绝域也相应的变化：当原假设是$\mu=0$时是双边检验，其他两种情况是单边检验。

R实战演习

在R中，配对样本检验与独立样本检验用同样的函数，只需在使用函数t.test()时设定paired=TRUE就可以了。

```
> before = c (94.5,101,110,103.5,97,88.5,96.5,101,104,116.5)
>  after = c (85,89.5,101.5,96,86,80.5,87,93.5,93,102)
> t.test (before,after,paired=T)
    Paired t-test
data:  before and after
t = 14.1641, df = 9, p-value = 1.854e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  8.276847 11.423153
sample estimates:
mean of the differences
      9.85
```

由输出结果可知，拒绝原假设，说明该健美俱乐部声称“参加其训练班至少可以使肥胖者平均体重减轻8.5kg以上”的说法还是有一定根据的。

7.4　比例假设检验

现实中有很多数据是比例数据，有时需要对比例数据进行假设检验。这一节主要介绍单样本的比例检验和两样本比例检验。

7.4.1　单样本比例检验

例7.9　同参数估计中一样，首先考虑一个简单的调查问题：为调查某大学男女比率是否是1：1，在校门处观察，发现100学生中有45个女性。那么，这是否支持该大学总体男性占比为50%的假设？

理论推导和检验步骤

设 $X \sim b(1,p)$，p 为事件发生的概率，$x_1,x_2,\ldots,x_n$ 是从总体X中抽取的样本，对比例 p 是否为 p_0 进行检验。

（1）检验假设。$H_0: p = p_0$ 和 H_1: $p \neq p_0$

（2）给出显著性水平 α。

（3）枢轴量：由于 $X \sim b(1,p)$，均值为 p，方差为 $p(p-1)$，当n较大时，根据中心极限定理，$\overline{x}$ 近似服从正态分布 $N(p, p(1-p)/n)$，枢轴量 $u = \dfrac{\overline{x}-p}{\sqrt{p(1-p)/n}}$。

在 H_0 成立时，统计量 $u = \dfrac{\overline{x}-p}{\sqrt{p(1-p)/n}} \sim N(0,1)$。

（4）计算 u 值对应的p值。

（5）若 u 值落在拒绝域内，则接受 H_1，反之不能拒绝 H_0。

当原假设变化时，拒绝域也相应的变化：当原假设是 $p = p_0$ 时，是双边检验，其他两种情况是单边检验。

R实战演习

本例中验证问题如下。

H_0：$p = 0.5$ 和 H_1：$p \neq 0.5$（这是一个双边假设检验）。

由于上述介绍的方法与软件中默认使用的函数prop.test()使用的检验不同，因此，在这里我们自行编写函数，然后使用上述介绍的方法进行假设检验。命令如下。

```
> proptest<-function(x,n,p,alternative)
{   Se=sqrt(p*(1-p)/n)
    u=(x/n-p)/Se
    if (alternative=="twoside")  p=2*(1-pnorm(abs(u)))
    else if (alternative=="less")  p=pnorm(u)
    else   p=1-pnorm(u)
    return(list(u=u, p=p))
}
> proptest(45,100,0.5,alternative="twoside")
$u
[1] -1
$p
[1] 0.3173105
```

注意，p值为0.3173，p值报告了我们原假设成立时的可能性大小。这里所谓的可能性大小，是相对于备择假设而言的。在这个例子中，备择假设是双边的，即检验统计量或者太小，或者太大。具体来说，p值在此例中为当有一半的人回答“是”时，被抽查到的人回答“是”的人数小于等于45或大于等于55的概率。

现在p值没有这么小，即通过本次观测，我们没有充分理由拒绝原假设，故接受原假设。

接下来，重复上面的例子，假设我们询问1000个人，有450人回答“是”，现在问原假设p=0.5是否还成立？

```
> proptest(450,1000,0.5,alternative="twoside")
$u
[1] -3.162278
$p
[1] 0.001565402
```

这次，p值比较小（0.001565），因此拒绝原假设。此例表明，p的取值不仅取决于比例，还和n有关。当n逐渐增大时，样本均值的标准误逐渐减少。简单地说，样本量越大，从样本获取的信息越充分，样本更能反映总体，我们统计推断得到的结论更加准确。

7.4.2　两样本比例检验

例7.10　两项调查：民意测验专家想知道某广告是否对观众产生明显影响，为此做了为期两周的调查，数据如下。

	第　一　周	第　二　周
喜欢	45	56
不喜欢	35	47

建立假设检验。H_0：$\pi_1=\pi_2$，双边备择假设H_1：$\pi_1 \neq \pi_2$。

在原假设下，假定两个总体比例是相等的（$\pi_1=\pi_2$），因此，总体比例的合并估计是基于原假设的，所以用合并两个样本相加成功数（x_1+x_2）除以总样本容量（n_1+n_2）。

构造枢轴量：

$$Z_{STAT}=\frac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim N(0,1)$$

其中：$\bar{p}=\frac{x_1+x_2}{n_1+n_2}$，$p_1=\frac{x_1}{n_1}$，$p_2=\frac{x_2}{n_2}$。

计算出Z的值，依照前面的U检验，做出判断。

R实战演习（软件与理论不同）

这里可使用命令prop.test去处理这类问题，我们只需知道何时和如何使用它即可。函数prop.test的用法为prop.test(x,n)，其中，x为实际观测数，n为总数。由于有两个x的取值，所以现在是验证两者喜欢的比率是否明显不同。输入如下命令：

```
>prop.test (c (45,56) ,c (45+35,56+47) )
2-sample test for equality of proportions with continuity correction
data: c (45, 56)  out of c (45 + 35, 56 + 47)
X-squared = 0.0108, df = 1, p-value = 0.9172
alternative hypothesis: two.sided
95 percent confidence interval:
-0.1374478 0.1750692
sample estimates:
prop 1      prop 2
0.5625000   0.5436893
```

观察可知，p值为0.9172，因此接受原假设，即$\pi_1=\pi_2$。

7.5 习题

1. 从一批钢管中抽取10根，测得其内直径（单位：mm）数据如下。

100.36 100.31 99.99 100.11 100.64
100.85 99.42 99.91 99.35 100.10

设这批钢管内直径服从正态分布$N(\mu,\sigma^2)$，试分别在下列条件下检验假设（$\alpha=0.05$）。

$$H_0: \mu=100 \text{ vs } H_1: \mu>100$$

（1）已知$\sigma=0.5$。

（2）σ未知。

2. 从小学五年级男学生中抽取20名，测量其身高（单位：厘米），其数据如下。

136 144 143 157 137 159 135 158 147 165
158 142 159 150 156 152 140 149 148 155

以$\alpha=0.05$做假设检验。

（1）H_0：$\mu=149 \text{ vs } H_1$：$\mu\neq100$

（2）H_0：$\sigma^2=75 \text{ vs } H_1$：$\sigma^2\neq75$

3. 某产铸造车间为提高铸件的耐磨性而试制了一种镍合金铸件以取代铜合金铸件，为此，从两种铸件中各抽取一个容量为8和9的样本，测得其硬度如下。

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

根据专业经验，硬度服从正态分布，且方差保持不变，试在显著性水平$\alpha = 0.05$下判断镍合金的硬度是否有显著提高？

4. 甲、乙两台机床加工某种零件，零件的直径服从正态分布，现从各自加工的零件中分别抽取7件和8件产品，测得其直径如下。

X（机床甲）：16.2 16.4 15.8 15.5 16.7 15.6 15.8

Y（机床乙）： 15.9 16.0 16.4 16.1 16.5 15.8 15.7 15.0

总体方差反应了加工精度，在显著性水平$\alpha = 0.05$下比较两条机床的加工精度有无差别？

5. 下面给出了两种型号的计算器充电以后所能使用的时间（h）的观测值。

型号A：5.5 5.6 6.3 4.6 5.3 5.0 6.2 5.8 5.1 5.2 5.9

型号B：3.8 4.3 4.2 4.0 4.9 4.5 5.2 4.8 4.5 3.9 3.7 4.6

试问能否认为型号A的计算器平均使用时间比型号B来得长（取$\alpha = 0.01$）？

6. 为了比较两种谷物种子的优劣，特选取10块土质不全相同的土地，并将每块土地分为面积相同的两部分，分别种植这两种种子，施肥与田间管理在20小块土地上都相同，下面是各小块上的单位产量：

土地	1	2	3	4	5	6	7	8	9	10
种子一的单位产量 x	23	35	29	42	39	29	37	34	35	28
种子二的单位产量 y	30	39	35	40	38	34	36	33	41	31

假定单位产量服从正态分布，试问：两种种子的平均单位产量在显著性水平$\alpha = 0.05$上有无差异？（成对数据检验）

7. 有一批蔬菜种子的平均发芽率$p_0 = 0.85$，现随机抽取500粒，用种衣剂进行浸种处理，结果有445粒发芽，试检验种衣剂对种子发芽率有无影响？

8. 据以往经验，新生儿染色体异常率一般为1%，某医院观察了当地400名新生儿，只有1例染色体异常，问该地区新生儿染色体异常是否低于一般水平？