假设检验

参数假设检验

主要涉及：
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、比例假设检验

假设检验的思想与步骤

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：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杯的已经调制好的茶，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，结果是女士准确的分辨出了10杯中的每一杯。

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

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0.510 ，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，而认为该女士有鉴赏能力。

若10杯中若只有6杯说对了，怎么判断这个问题呢？

若100杯中有60杯说对了，你怎么判断这个问题呢？

表：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即原假设和备择假设。先假定原假设成立，并在原假设成立条件下，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，我们都不能拒绝原假设，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。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。

假设检验的基本步骤

例：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，标准差的确是1.00cm。你作为该公司的检验员抽取样本，如何判定均值就是5.00cm呢？
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

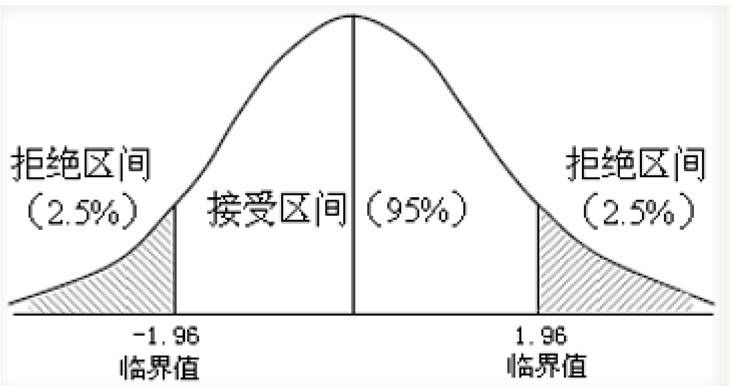
这是假设检验问题，即判断某一假定是否正确。

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：$\mu=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$，根据中心极限定理，可以证明在$H_0$成立时，$\mu:N(0,1)$。所以$\mu$落在如图中阴影覆盖部分是小概率时间，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$\alpha=0.05$.若$-1.96<\mu<1.96$,说明落在接受域，则不能拒绝H_0，若$\mu<-1.96$或者$\mu>1.96$，则拒绝原假设H_0, 接受H_1.



关于该检验，可以编写一个做两边检验的函数u.test进行检验。

In [ ]:

```
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统计量是0.5657252，对应的p值是0.5715806，即是说在H_0成立的前提下，u统计量 为0.5657252的概率为0.5715806。不能拒绝原假设。

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和备择假设。注意这两个的假设互换可能会导致结论 有差异。

2、找到合适的枢轴量。

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。

显著性水平如何选取？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。容易知道，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，只能增加样本量，在现实生活中增加样本量有时是可行的，有时是不可行的。

显著性水平的大小要根据具体研究问题来选取。在很多统计软件中，为了避免显著性水平大小的选取给出u值相应的P值，让读者自己取舍。

正态总体单样本参数假设检验

均值的检验

（1）方差已知情形

例：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。现在问，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：

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

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

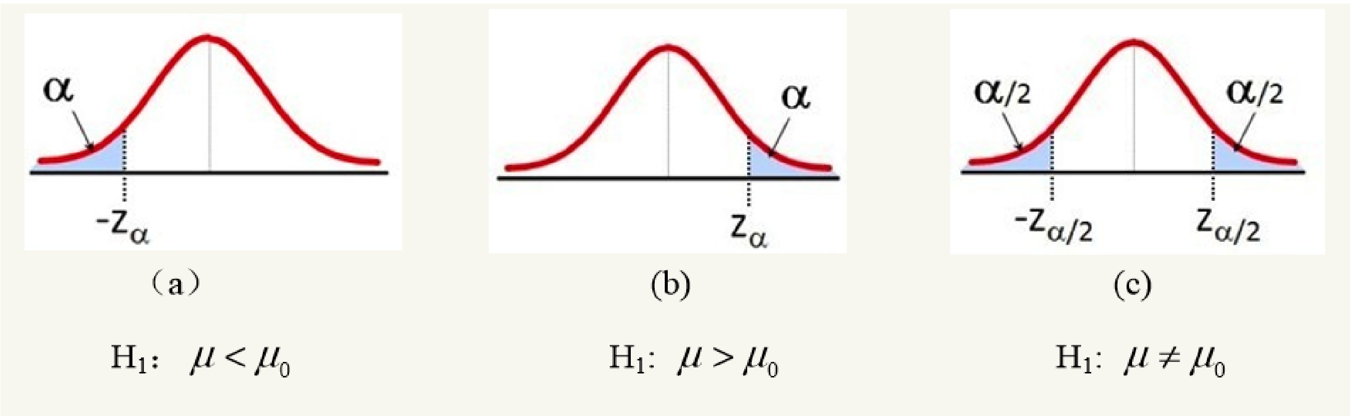
（2）给定检验水平α

（3）枢轴量$u = \frac{\overline{x}-H_0}{\sigma/\sqrt{n}}$服从标准正态分布。

（4）计算u统计量对应的p值。

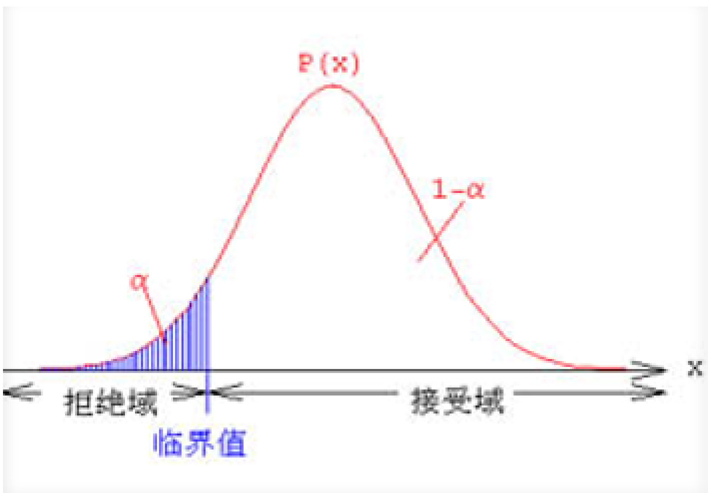
（5）u值落在拒绝域内，则拒绝H_0,反之不能拒绝H_0.

容易推导，若备择假设变化时，拒绝域的变化如图所示。通常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：$u = \frac{\overline{x}-H_0}{\sigma/\sqrt{n}}$。在$\mu = \mu_0$时，$\mu$服从$N(0, 1)$。也就是说在原假设成立条件下，$\mu$落在下图拒绝区域中的概率会比较小。通过计算$\mu$值对应的$p$值。若$p \leq \alpha$，拒绝$H_0$,接受$H_1$。若$p > \alpha$，接受$H_0$,拒绝$H_1$.



前面编写u.test函数进行双边检验，此处在前文编写的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。

In [ ]:

```
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
```

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：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，如果新的资产公司平均收益率大于原来资产管理公司的平均收益率，公司将选择新的资产管理公司。原来的资产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，问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$x_1,x_2,\ldots\ldots,x_n$来自$X:N(\mu,\sigma^2),\sigma^2$未知，对均值$\mu$进行检验。

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

（2）给定检验水平α

（3）枢轴量$t=\frac{\overline{x}-H_0}{s/\sqrt{n}}$服从标准正态分布。

（4）计算t统计量对应的p值。

（5）t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，拒绝域的变化同上:当原假设是$\mu=\mu_0$时候是双边检验，其他两种情况是单边检验。

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

下面用R的内部函数t.test进行检验：

In [ ]:

```
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")
```

检验的p值=0.0065<0.05，在显著性水平α=0.05时拒绝原假设，接受备择假设，可认为新资产管理公司的收益率的确更高.

方差检验

方差的大小表现的是总体的离散程度，在很多现实情况需要对方差进行检验。

例：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,均值和 标准差单位是微克。通过抽查20个废水样品，如何判废水处理设备正常工作？数据见R程序中 变量x。对于这个问题，需要检验均值是否是500ug和标准差是否20ug。这两个指标若有差异， 都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。关于均值的检验在上一节已经介绍过了。本节将讨论标准差的检验。

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

（1）检验假设：$H_0:\sigma^2=\sigma_0^2,H_1:\sigma^2\neq\sigma_0^2$

（2）给定检验水平α

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

（4）计算χ^2统计量对应的p值。

（5）χ^2值落在拒绝域内，则接受H_1,反之不能拒绝H_0.

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

In [ ]:

```
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)
```

通过p值可以知道，不能说明机器工作不正常，也就是默认机器正常工作。

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。本节由于均值的检验中 需要用到方差的检验，所以本节先介绍方差的检验，然后均值检验。

双样本方差的检验（方差齐性检验）

例： 假设你是一个轮胎生产商，你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$X_1:N(\mu_1,\sigma_1^2)$和$X_2:N(\mu_2,\sigma_2^2)$，请问你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？

理论推导和检验步骤（F检验）

用数学语言描述，总体$X_1:N(\mu_1,\sigma_1^2)$和$X_2:N(\mu_2,\sigma_2^2)$，X_1和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σ_1^2与σ_2^2是否相等，或者说哪个明显的大？

对σ_1^2是否与σ_2^2相等进行检验：

（1）检验假设：$H_0:\sigma_1^2=\sigma_2^2, H_1:\sigma_1^2\neq\sigma_2^2$

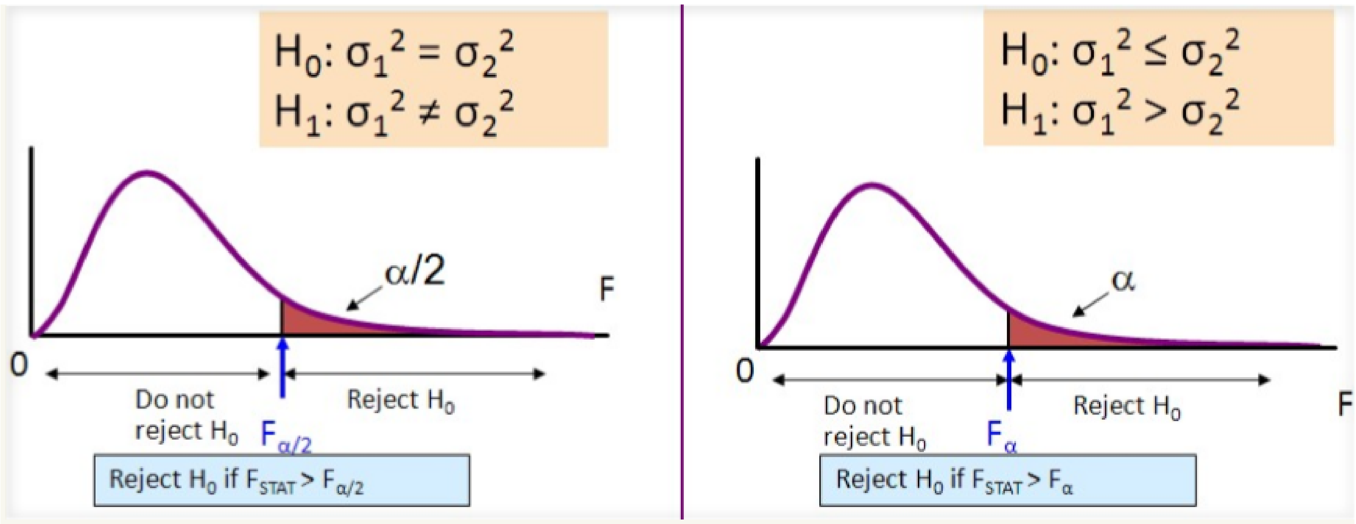
（2）给定检验水平α

（3）枢轴量$F=\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$服从$F(n_1-1,n_2-1)$。在$H_0$成立时，统计量$F=\frac{s_1^2}{s_2^2}\sim F(n_1-1,n_2-1)$。

（4）计算F值对应的p值。

（5）F值落在拒绝域内，则接受H_1,反之不能拒绝H_0.

同上推导相似，若原假设变化时，拒绝域的变化同上： 当原假设是$\sigma_1^2=\sigma_2^2$时候是双边检验，其他两种情况是单边检验。‘



7-4 不同原假设下的检验拒绝域

In [ ]:

```
#x1=floor(rnorm(20,38,9))
#x2=floor(norm(20,38,4))
#var.test(x1,x2)
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)
```

在显著性水平0.05的条件下，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。

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，但在计算上则有一些区别。

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，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的平 均时间快吗？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

要具体检验以下假设：

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

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$x_1,x_2,\ldots\ldots x_n$来自总体$X:N(\mu_1,\sigma^2)$，$y_1,y_2,\ldots\ldots y_m$来自总体$Y:N(\mu_2,\sigma^2)$，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）给定检验水平α

（3）枢轴量$t=\frac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{s_w/\sqrt{\frac{1}{m}+\frac{1}{n}}}$，其中$s_w=\frac{1}{m+n-2}[\sum_i^m=1x_i-\overline{x}+\sum_i^n=1y_i-\overline{y}]$，在$H_0$成立时，统计量$t=\frac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{s_w/\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2)$。

（4）计算t统计量对应的p值。

（5）t值落在拒绝域内，则拒绝H_0,反之不能拒绝H_0.

若原假设变化时，拒绝域的变化同上:当原假设是$\mu_1=\mu_2$时候是双边检验，其他两种情况是单边检验。

2、方差不齐性时

方差不齐性的情况下理论推导和检验步骤：

方差不齐时，设$x_1,x_2,\ldots\ldots x_n$来自总体$X:N(\mu_1,\sigma^2)$，$y_1,y_2,\ldots\ldots y_m$来自总体$Y:N(\mu_2,\sigma^2)$，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）给定检验水平α

（3）枢轴量$t=\frac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}}$，其中$s_x^2$为来自总体$X$的样本的样本方差，$s_y^2$为来自总体$Y$的样本的样本方差，在$H_0$成立时，统计量$t=\frac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}}\sim t(l)$。其中$l=(\frac{s_1^2}{m}+\frac{s_2^2}{n})^2/(\frac{s_1^4}{m^2(m-1)}+\frac{s_2^4}{n^2(n-1)})$。

$s_d=[\frac{1}{n-1}\sum_i^n=1(d_i-\overline{d})^2]^{\frac{1}{2}}$。在$H_0$成立时，统计量$t=\overline{d}/(s_d/\sqrt{n})\sim t(n-1)$

（4）计算t统计量对应的p值。

（5）t值落在拒绝域内，则拒绝H_0,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$\mu=0$时候是双边检验，其他两种情况是单边检验。

Step1在作两样本均值检验时，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。）

Step2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（上面已经介绍）。

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)

p = 0.0379<0.05，说明两组数据的方差是不一样的。

Step3 t检验判断均值。对于R软件中，两个独立样本检验两种情况都用一个函数，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。

In [ ]:

```
t.test(x1,x2)
```

经检验，p值=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，而不是假定他们是独立正态分布的。

基本的模型是$Y_i = X_i + \epsilon_i$，其中ϵ_i为随机项。我们想检验ϵ_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：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

要具体检验以下假设：

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

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

对均值μ是否为0进行检验：

（1）检验假设：$H_0: \mu=0, H_1: \mu \neq 0$

（2）给定检验水平α

（3）枢轴量$t=\frac{\overline{d}}{sd/\sqrt{n}}$，其中$d_i=x_i-y_i, \overline{d}=\frac{1}{n}\sum_i^n=1d_i$服从标准正态分布。

（4）计算t统计量对应的p值。

（5）t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，拒绝域的变化同上:当原假设是$\mu=\mu_0$时候是双边检验，其他两种情况是单边检验。

在R中，配对样本检验与独立样本检验是用同样的函数，只要在使用函数t.tes t时设定paired=TRUE就可以了。

In [ ]:

```
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)
```

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。

比例假设检验

主要介绍单样本的比例检验和两样本比例检验。

单样本比例检验

例: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，$x_1, x_2.\dots\dots x_n$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：$H_0: p = p_0; H_1: p \neq p_0$

(2)给出显著性水平α

(3)枢轴量：由于$X: b(1,p)$,均值为p,方差为$p(p-1)$，当n为比较大时，根据中心极限定理，$\overline{x}$近似服从正态分布$N(p, p(p-1)/n)$，枢轴量$u = \frac{\overline{x}-p}{\sqrt{p(p-1)/n}}$。在$H_0$成立时，统计量$u = \frac{\overline{x}-p}{\sqrt{p(p-1)/n}} \sim N(0,1)$

(4)计算u值对应的p值。

(5)u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，拒绝域也相应地变化：当原假设是$p = p_0$时候是双边检验，其它两种情况是单边检验。

本例中验证问题如下：

$$H_0: p = 0.5; H_1: p \neq 0.5$$

由于上述介绍的方法与软件中默认使用的函数prop.test使用的检验有区别，在这里我们自行编写函数进行使用上述介绍的方法进行假设检验。命令如下：

In [ ]:

```
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")
```

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=0.5是否还成立？

In [ ]:

```
proptest(450,1000,0.5,alternative="twoside")
```

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， 还和样本量有关。特别地，当样本量逐渐增大时，样本均值的标准误逐渐减少。即样本量 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。

两样本比例检验

例: 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：

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

其中：$\overline{p}=\frac{X_1+X_2}{n_1+n_2},p_1=\frac{X_1}{n_1},p_2=\frac{X_2}{n_2}$。

计算出Z的值，依照前面的U检验，做出判断。

这里可使用命令prop.test去处理此种问题。我们只需知道何时和怎样使用它。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，现在是验证两者喜欢的比率是否明显的不同。输入如下命令：

In [ ]:

```
prop.test(c(45,56),c(45+35,56+47))
```

由p值=0.9172可知，接受原假设。

非参数假设检验

图示法

例：验证每加仑汽车里程数的总体分布类型美国1974年《MotorTrendUS》杂志给出了一项调查数据，关于汽车燃油消耗和十项 汽车设计和性能。R软件将其放入mtcars数据集中。其中mpg变量是汽车每加仑汽油的行驶里程。现在请问汽车每加仑行驶里程服从什么分布？

21.0 18.7 22.8 17.3 14.7 21.5 19.2 15.8

21.0 18.1 19.2 15.2 32.4 15.5 27.3 19.7

22.8 14.3 17.8 10.4 30.4 15.2 26.0 15.0

21.4 24.4 16.4 10.4 33.9 13.3 30.4 21.4

在没有其他任何信息以前，可以通过图示的方法直观地分析。这个我们在第6章的探索性分析有详细地讲解。此处我们仅作简单介绍.

(1)直方图(histogram)

In [ ]:

```
head(mtcars)
attach(mtcars)
hist(mtcars$mpg,prob=T,col="lightblue") #绘制直方图
xfit<-seq(min(mtcars$mpg),max(mtcars$mpg),length=40)
yfit<-dnorm(xfit,mean(mtcars$mpg),sd(mtcars$mpg)) #你猜测的随机变量的密度分布
lines(xfit,yfit,col="red",lwd=3) #画出密度函数，与直方图对比，直观判断
```

(2)茎叶图(Stem-and-LeafDiagrams)

In [ ]:

```
stem(mtcars$mpg)
```

特殊的，若验证样本来源于正态总体，称之为正态性检验。除了直方图，还有QQ图等可 以验证样本是否来源于正态总体。

(3)Q-Q图：

仅用于验证正态性以标准正态分布的分位数为横坐标，以处在相同百分位的样本分位数为纵坐标，把样本 表现在直角坐标系中的散点。如果资料服从正态分布，则样本点应该呈一条围绕第一象限对角线的直线。

也可以用累计概率来作图，成为P-P，但相较之下以Q-Q图为佳，效率较高。

In [ ]:

```
attach(mtcars)
qqnorm(mtcars$mpg)
qqline(mtcars$mpg)
```

虽然耗油量大了以后，离得比较远，可以大体认为比较靠近直线。其他图示法还有经验分布图、箱式图(观测离群值和中位数)等方法。下面将介绍量化检验方法。

卡方检验

卡方检验是通过构造卡方枢轴量，判别是否落入拒绝域来判别是否接受原假设。首先简单介绍一下卡方分布。

卡方分布

若Z_i独立同分布于$N(0,1)$，令$\chi^2=\sum_{i=1}^{n}Z_i^2$，那么$\chi^2$服从自由度为$n$的卡方分布。

通过R编写程序可以画出不同自由度(df)下的卡方分布的密度函数曲线图。

x=seq(0,20,0.1)#生成x序列 curve(dchisq(x,2),0,20,ylab="p(x)") curve(dchisq(x,4),add=T,lty=2) curve(dchisq(x,6),add=T,lty=3) curve(dchisq(x,8),add=T,lty=4) curve(dchisq(x,10),add=T,lty=5) legend(13,0.4,c("df=2","df=4","df=6","df=8","df=10"),lty=1:5,bty="n")

从图我们可以看到，卡方分布的形状取决于其自由度df。

R中生成卡方分布随机数的函数是rchisq() ，图8-4是生成了自由度为5时的100个卡方随 机数据的探索性图示分析。

In [ ]:

```
X5=rchisq(100,5)
EDA<-function(x)
{
par(mfrow=c(2,2))#同时做4个图（设置作图窗口为2行2列格式）
hist(x);#直方图
dotchart(x);#点图
boxplot(x,horizontal=T);#箱式图，详见6.2.2
qqnorm(x);
qqline(x)#正态概率图
par(mfrow=c(1,1))#恢复成单图
}
EDA(X5)
```

由图可知自由度较小时的卡方分布偏 态比较严重，然而当自由度逐渐变大时分布将 趋于正态(这点也可以从图看出)。

卡方拟合优度检验

卡方拟合优度检验(Chi-squared goodness of fit tests)用来检验样本是否来自于特定类 型分布的一种假设检验。下面分离散型和连续性随机变量介绍该检验。

(1)、离散型分布验证

例:骰子是均匀的吗? 如果掷一骰子150次并得到以下分布数据，此骰子是均匀的吗？

点数 1 2 3 4 5 6 合计

出现次数 22 21 22 27 22 36 150

若骰子是均匀的，你会理所当然地认为各面出现的概率都一样(1/6)，即在150次投掷中 骰子的每一面将期望出现25次。但表中数据表明数字为6的一面却出现了36次，这是纯属巧合还是其它什么原因？

回答这个问题的关键是看观测值与期望值离的有多远。如果令f_i为观测到的第i类数据的出现频数，e_i为第i类数据出现次数的理论期望值，则χ^2统计量可表示为：

$$\chi^2 = \sum_{i=1}^{n} \frac{(f - e_i)^2}{e_i}$$

直观地看，如果实际观测频数和理论预期频数相差很大，卡方统计量的值将会很大；反 之则较小。同时，数据必须为独立同分布的。如果这些假定都满足，那么卡方统计量将近似 服从于自由度为n-1的卡方分布。建立假设检验，原假设为各面出现的概率为理论值，备择假设为六个面中的一些或全部出现概率不等于理论值。

下面进行检验，R有针对这类问题的内部函数，在使用它们之前首先要指定实际频数和理 论预期概率。

在此例中，用法很简单：

In [ ]:

```
freq=c(22,21,22,27,22,36)
probs=c(1,1,1,1,1,1)/6#指定理论概率(多项分布)
chisq.test(freq,p=probs)
```

前面所述假设检验假定，原假设为第i类对应的概率为pi (在此例中pi =1/6)，备择假设为 至少有一类对应的概率不等于pi 。我们看到，卡方值为6.72，自由度为df=6-1=5。p值为0.2423，所以没有理由拒绝骰子是均匀的假设。

例: 字母的分布：英文字母5个中常用的英文字母近似地服从下述分布(真实的分布其 实是关于全部26个字母的，现简化为只有5个字母的情况)：

letter E T N R O

freq 29 21 17 17 16

上述分布的意思是，当字母E、T、N、R、O出现时，平均100次中有29次是字母E而不 是其它4个字母，此信息对在密码学中破译一些基本的密码时非常有用。假设分析一篇文章 ，计算字母E、T、N、R和O的出现次数，得到如下频数分布：

letter E T N R O

freq 100 110 80 55 14

为知道字母出现次数的比例是否依次为0.29，0.21，0.17，0.17，0.16，做一个卡方拟合优度检验。

In [ ]:

```
freq=c(100,110,80,55,14)
probs=c(29,21,17,17,16)/100
chisq.test(freq,p=probs)
```

检验的p-value=2.685e-11，拒绝原假设，这表明文章不大可能是用英语写的。

(2)、连续型分布验证

对于连续型分布的验证，本质是将其离散化，即分成相应的区域，通过每个区间的理论 概率及频数构造枢轴量。

具体如下：原假设样本来源于某特定分布，将数轴$(-\lim,+\lim)$分成k个区间：$I_1=(-\lim,a_1),I_2=[a_1,a_2),\ldots,I_k=[a_{k-1},+\lim)$，记这些区间的理论概率分布为$p_1,p_2,\ldots,p_k,p_i=P\{X\in I_i\},i=1,2...,K$。$f_i$为$X_1,X_3,\ldots,X_n$中落在区间$I_i$内的个数，则同上构造枢轴量：$\chi^2=\sum_{i=1}^k\frac{(f_i-np_i)^2}{np_i}$。在原假设成立的条件下，当$n\to\lim$时，$\chi^2$依分布收敛于自由度为$k-1$的$\chi^2$分布。当给定显著性水平$\alpha$，$\chi^2>\chi^2_\alpha(k-1)$，则拒绝原假设。同样也可以通过$p-value$来判断。

当连续型分布中有未知参数时，先估计出未知参数，同样的方法构造枢轴量，但其服从的卡方分布自由度将会降低r，r是被估参数的个数。上述过程写成数学语言如下：

参数已知时：当$n\to\lim,\chi^2=\sum_{i=1}^k\frac{n}{p_i}(\frac{f_i}{n}-p_i)^2=\sum_{i=1}^k\frac{(f_i-np_i)^2}{np_i}\chi(k-1)$

含有r个未知参数：当$n\to\lim,\chi^2=\sum_{i=1}^k\frac{n}{\hat{p}_i}(\frac{f_i}{n}-\hat{p}_i)^2=\sum_{i=1}^k\frac{(f_i-n\hat{p}_i)^2}{n\hat{p}_i}\chi(k-r-1)$

若原假设为真时，χ^2应较小，否则就怀疑原假设，从而拒绝域为$R=\{\chi^2\geq d\}$，对于给定的α有$P\{\chi^2\geq d\}=\alpha$，即d是卡方分布的α上分位数。

例: 验证男性身高是否服从N(170，64) 假如要检验成人男性是否服从分布，调查了20名男性的身高为

159.8 178.5 168.9 183.2 174.0 160.9 180.0 171.7 152.4 174.3

170.2 185.3 169.6 160.1 158.9 164.6 172.2 168.0 182.1 171.1

请问男性身高是否服从N(170，64)分布？

In [ ]:

```
x<-c(159.8,178.5,168.9,183.2,174.0,160.9,180.0,
    171.7,152.4,174.3,170.2,185.3,169.6,160.1,158.9,164.6,172.2,182.1,171.1)
fn=table(cut(x,breaks=c(min(x),160,170,180,190,max(x))))
F=pnorm(c(min(x),160,170,180,190,max(x)),170,8)
P=c(F[1],F[2]-F[1],F[3]-F[2],F[4]-F[3],1-F[4])
chisq.test(fn,p=P)
```

注意：卡方拟合优度检验的检验结果依赖于分组，当不能拒绝原假设时，并不能说明原假设 就是正确的，特别是数据量不够大的时候。而且卡方检验是把连续数据转换成离散数据后进行检验 ，在转换过程中存在信息丢失，所以对卡方分布检验的结果需要小心。

卡方独立性检验

卡方独立性检验(Chi-squaredtestsofindependence)，在原假设两个因素相互独立的 前提下，比较两个及两个以上样本率(构成比）以及两个分类变量的关联性分析。其根本思想 就是在于比较理论频数和实际频数的吻合程度或拟合优度问题。基于这一原理，构造同前章 节中的卡方统计量来检验列联表中的两个因子是否相互独立。

例8.5、安全带与受伤程度的独立性检验 假若你得到如下列表数据：在遇到车祸的情况下，乘客有系安全带和没系安全带时受到的冲 击力。

受伤情况 无 轻微 较重 严重

系安全带 12813 647 359 42

没系安全带 65963 4000 2642 303

各因子之间是否独立？安全带是否起作用？对此我们可使用卡方检验。但是，理论预期频数为多少？在独立性的原假设下，P(系安全带无受伤)= P(系安全带)P(无受伤)是由回答“否”（各列之和除以n）的比例和回答“是”（各行之和除以n）的比例来估计的。那么此单元格的预期频数即为上式乘以n；或者简单地，各行之和乘以各列之和再除以n，即$\frac{n1}{n}\cdot\frac{n2}{n}\cdot n$。由于需要对每个单元格都这样做，所以最好由计算机来完成。

In [ ]:

```
yesbelt=c(12813,647,359,42)
nobelt=c(65963,4000,2642,303)
chisq.test(rbind(yesbelt,nobelt))
```

上述过程检验了两行因子相互独立的原假设，极小的p值说明应该拒绝原假设，即它们不 是相互独立的。

卡方两样本同质性检验

卡方同质性检验(Chi-squaredtestsforhomogeneity)，检验各行是否来自同一个总体 。

直观地，如果各行因子来自于相同的总体，每一类的出现概率应该是差不多的，而卡方统 计量则将再次帮助我们解释“差不多”的含义。

例:两样本的总体分布是否有差异我们知道，卡方同质性检验用于检验分类数据的各行因子是否来自不同的总体。其检验的有效性如何？

通过使用sample命令，我们可以很容易地模拟抛掷一颗骰子。抛一个均匀的 ，再抛一个不均匀的，以测试卡方检验能否检验出它们的差别。

首先抛掷均匀骰子100次，再抛掷不均匀骰子100次，列表如下：

die.fair=sample(1:6,100,p=c(1,1,1,1,1,1)/6,rep=T)#均匀骰子
die.bias=sample(1:6,100,p=c(.5,.5,1,1,1,2)/6,rep=T)#不均匀骰子
res.fair=table(die.fair);res.bias=table(die.bias) count=rbind(res.fair,res.bias) count

卡方同质性检验所做的分析和卡方独立性检验是一样的：对于每一个单元格它计算出预期频数，然后使用预期频数与观测到的频数进行比较。那么，理论预期频数应为多少？

考虑对于一个均匀的骰子，出现数字2的次数会有多少？预期频数应为100乘以出现数字2的概率。 这个我们并不知道，但如果假定两个行因子服从相同的分布，那么，边际概率将给出一个 估计。总的边际值为25/200=(21+4)/(100+100)，所以预期得到100*(25/200)=12.5，而 实际上我们得到的频数是21。

我们将这些离差平方除以预期值并加总起来，得到统计量：

$$\chi^2 = \sum \frac{(f_i - e_i)^2}{e_i}$$

在原假设下，两个数据集都来自于相同的总体(同质性)，且样本服从于自由度为(2-1)(6-1)=5(即行数减1乘以列数减1)的卡方分布。以上过程可由函数chisq.test实现：

In [ ]:

```
chisq.test(count)
```

注意到p值很小，充分说明它们的非同质性。如果想要输出预期值，可取检验结果中 的expected值：

In [ ]:

```
chisq.test(count)$exp#获得期望频数
```

秩和检验

秩和检验则是一种基于计算秩和，在没有样本先验信息前提下做出的检验。

秩的概念

非参数检验通常是将数据转换成秩来进行分析的。

一般来说，秩就是该数据按照升幂排 列之后，每个观测值的位置。

例:对保险数据编秩：

In [ ]:

```
x=c(21240,4632,22836,5484,5052,5064,6972,7596,14760,15012,18720,9480,4728,67200,52788)
(Ri=rank(x))
```

利用秩的大小进行推断就避免了不知道数据分布的困难。这也是大多数非参数检验的优点 。多数非参数检验明显地或隐含地利用了秩的性质。

单样本符合秩检验

前面已经知道，车险索赔额不服从正态分布，所以对其作检验就不能用t检验，而需要用非参数的wilcoxon符号秩检验，用R做此检验可使用函数wilcox.test。

例8.8在上面关于保险公司车险索赔的调查中，已知上年车险索赔的中位数为5080元， 问是否可以说当年的索赔额的中位数与上年有显著的不同？

首先，绘制一个茎叶图：

x=c(21240,4632,22836,5484,5052,5064,6972,7596,14760,15012,18720,9480,4728,67200,52788) stem(x)

由茎叶图可看出分布偏态并呈拖尾，于是排除t检验而考虑中位数检验。

原假设为中位数等于5080，备择假设为中位数不等于5080。

In [ ]:

```
wilcox.test(x,mu=5080)
```

符号秩检验wilcox.test可以看成一种中位数检验，许多书也介绍符号检验，即只要考虑 比中位数大还是小的符号，而不考虑秩。

也可以编写一个R函数median.test来完成，它将计算中位数双边检验的p值。

In [ ]:

```
median.test<-function(x,median=NA)
{
x<-as.vector(x)
n<-length(x)
bigger<-sum(x>median)
equal<-sum(x==median)
count<-bigger+equal/2
count<-min(count,n-count)
p<-2*pbinom(count,n,0.5)
c(Positive=bigger,Negative=count,P=p)
}
median.test(x,median=5080)
```

这里Positive表示正号有11个，Negative表示负号有4个，p值为0.1185，结果不显著。

注意：符号检验只考虑比中位数大还是小的符号，相对于wilcoxon符号秩检验，其利用的信息量较少，因此，我们推荐使用wilcoxon符号秩检验。

两独立秩和检验

两样本wilcoxon秩和检验也可由函数wilcox.test完成，其本质是一种非参数的检验方法 ，用法和单样本检验相似。

假定第一个样本有m个观测值，第二个有n个观测值。把两个样本混合之后把这m+n个观测值按升幂排序，记下每个观测值在混合排序下面的秩。再分别把两个样本所得到的秩相加。

记第一个样本观测值的秩的和为W1，而第二个样本秩的和为W2。这两个值可以互相推算，称为Wilcoxon统计量。该统计量的分布和两个总体分布无关。由此分布可以得到p值。

直观上看，如果W1与W2之中有一个显著地大，则可以选择拒绝原假设。

该检验需要的唯一假定就是两个总体的分布 有类似的形状，但不一定要求分布是对称的。

多个独立样本的秩和检验

多个独立样本的非参数检验是用Kruskal-Wallis秩和检验，这个检验的目的是看多个总体的位置参数是否一样。一般提法如下：设有k个连续型随机变量总体$X_1, X_2, \ldots, X_k$。$x_{i1}, x_{i2}, \ldots, x_{in_i}$是来自第$i$个总体$X_i$的样本，其容量为$n_i$，$i = 1, 2 \ldots k$.总的样本容量为$N = \sum_{i=1}^{k} n_i$。所有$N$个样本单元都是相互独立的。设第$i$个总体$X_i$的分布函数为$F(x - \theta_i), i = 1, 2 \ldots k$。这也就是说，这$k$个随机变量总体$X_1, X_2, \ldots, X_k$的分布函数的形状全都相同，仅有可能是位置参数不同。Kruskal-Wallis检验用于这k个位置参数$\theta_1, \theta_2, \ldots, \theta_k$是否全都相等的检验问题。它的原假设和备择假设分别为：

$H_0 : \theta_1 = \theta_2 = \ldots = \theta_k, H_1 : \theta_1, \theta_2, \ldots, \theta_k$不全都相等

显然，在原假设为真时，这k个连续型随机变量总体$X_1, X_2, \ldots, X_k$同分布。

在方差相等的正态总体假设下，这k个位置参数$\theta_1, \theta_2, \ldots, \theta_k$可理解为$k$个正态总体的均值，可使用ANOVA方法检验是否相等。

非参数的检验方法如下：

为简化讨论，不妨假设所有N个样本单元互不相等。在原假设成立，即$\theta_1 = \theta_2 = \ldots = \theta_k$时，所有$N$个样本单元都是独立同分布的。为此我们将这$k$组样本合在一起，计算每一个数据在合样本中的秩。记$x_{ij}$在合样本$\{x_{i1}, x_{i2}, \ldots, x_{in_i}, i = 1, 2, \ldots, k\}$中的秩为$R_{ij}$，$R_{ij} = 1, 2, \ldots, N$。Kruskal-Wallis检验的基本思想是用$x_{ij}$的秩$R_{ij}$代替$x_{ij}$，然后用ANOVA方法作统计分析。

秩的总的平均为：$\overline{R} = \sum_{i=1}^{k} \sum_{j=1}^{ni} R_{ij}/N = (N+1)/2$，

总平方和为常数：$SST = \sum_{i=1}^{k} \sum_{j=1}^{ni} (R_{ij} - \overline{R})^2 = \sum_{i=1}^{k} \sum_{j=1}^{ni} R_{ij}^2 - N\overline{R}^2 = N(N^2-1)/12$

组间平方和SSB为：$SSB = \sum_{i=1}^{k} n_i(\overline{R_i} - \overline{R})^2 = \sum_{i=1}^{k} n_i(\overline{R_i} - (N+1)/2)^2$

组内平方和SSW为：$SSW = \sum_{i=1}^{k} \sum_{j=1}^{ni} (R_{ij} - \overline{R_i})^2$

其中$\overline{R_i} = \sum_{j=1}^{ni} R_{ij}/n_j$是来自第{i}个总体的样本的秩的平均，$i = 1, 2, .., k$。由$SSB + SSW = SST$知，如果算的组间平方和$SSB$的值，那么也就得到了它的组内平方和$SSW$的值：所以基于秩的ANOVA仅需要计算$SSB$的值，且算的$SSB$的值于计量单位无关，这个$SSB$就可以作为检验统计量。

我们在SSB的值比较大的时候认为这k个位置参数不全都相等。W.H.Kruskal和W.A.Wallis将SSB乘以成熟因子$12/N(N+1)$，从而将SSB变换为$\frac{12}{N(N+1)}SSB = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i(\overline{R_i} - (N+1)/2)^2 = \frac{12}{N(N+1)} \sum_{i=1}^{k} R_{i+}^2/n_i - 3(N+1)$。显然，我们在$H$的值比较大的时候认为这$k$个位置参数不全都相等。 H作为检验统计量的检验称为Kruskal-Wallis检验。其中$R_{i+} = n_i\overline{R_i} = \sum_{j=1}^{ni} R_{ij}$是来自第$i$个总体的样本的秩和，$i = 1, 2...k$。

可以证明，在原假设H_0为真，即k个总体$X_1, X_2, \ldots, X_k$为同一连续型分布时，若$min\{n_1, \ldots, n_k\} \rightarrow \lim$，且对所有的$i = 1, 2, \ldots, k$都有$n_i/N \rightarrow \lambda_i \in (0, 1)$，则Kruskal-Wallis检验统计量$H$渐进服从$\chi^(k-1)$分布。这里要主要的事若在有相同观测值时对秩取平均，则统计量也要做相应修正。

Kruskal-Wallis检验为非参数检验，可在数据非正态的情况下代替单因素方差分析。

它的使用方式和用Wilcoxen符号秩检验代替t检验是一样的。另外它也是一个基于原始数据秩的 检验，故不要求数据的正态性。

当你不肯定单因素检验中的正态性假定是否成立时，便可考虑Kruskal-Wallis检验。它在R中的用法和oneway.test相似：

例：对R自带的植物生长数据进行Kruskal-Wallis检验。

In [ ]:

```
data(PlantGrowth)
head(PlantGrowth)
```

In [ ]:

```
kruskal.test(weight~group,data=PlantGrowth)
```

也可以将它作为一个数据框来直接调用，即kruskal.test(dtatframe)。注意到结果中的p 值虽然很小，但却比方差分析中的p值要大。不过，两种检验都表明原假设值得怀疑。

例：奖学金评级者的主观差异性验证假定学校要对300份奖学金申请作(做)出评价。若只有一个人做此项工作，工作量极大，故(因此)安排6个评价人。奖学金委员会希望确保每位评价人以相同的标准去评价每一份申请表，否则学生们得到的对待将是不公正的。检查各评价人是否公正的一种方法，是随机地安排50份申请表给各评价人进行评价，然 后比较6个评价人的评分。如果评价人的标准是一致的，那么可认为他们做出的评价差异是由随机误差所致。通常，评分等级并不服从正态假定，所以这类问题一般要用非参数方法进行 检验。

假如我们抽取三个评审人共36份申请表，并假定评分为5等级制，评分为5则意味着好 ，评分为1意味着差。数据如下：

评价者1：4，3，4，5，2，3，4，5，4，4，5，4

评价者2：4，4，5，5，4，5，4，4，5，5，4，5

评价者3：3，4，2，4，4，5，3，4，2，2，1，1

现在需要我们检验三个评审人的评价标准是否一致？ 将数据以如下方式输入至R中，并构造一个数据框：

In [ ]:

```
scores=c(4,3,4,5,2,3,4,5,4,4,5,4,4,4,5,5,4,5,4,4,5,5,4,5,3,4,2,4,4,5,3,4,2,2,1,1)
person=c(1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3,3,2)
boxplot(scores~person)#做一个箱线图比较分布
```

从图中我们可看出它们的分布明显的不同，并且评价者2 和评价者1、3分布有明显差异。Kruskal-Wallis检验允许我 们检验是否所有评价者的评分情况一样。

In [ ]:

```
kruskal.test(scores~person)
```

我们看到p值为0.3793，这意味着拒绝均值相等的原假设。

多个相关样本的秩和检验

多个相关样本的秩和检验是用Friedman方法，一般提法如下。设有一个k个处理b个区间的区组设计。假设第i个处理在第j个区间的观察值x_ij的分布函数为$F_j(x-\theta_i), i=1,2,\ldots,k, j=1,2,\ldots,b$。这也就是说，同一个区组，例如第$j$个区组内的$k$个观察值$x_{1j},x_{2j},\ldots,x_{kj}$的分布函数的形状都相同，仅有可能的是位置参数不同。而不同区组内的观察值的分布函数的形状有可能不同。k个处理b个区组的区组设计共有$N=bk$个观察值。所有这N个观察值都是相互独立的。Friedman检验用于这k个位置参数$\theta_1,\theta_2,\ldots,\theta_k$是否全都相等检验问题。它的原假设和备择假设分布为:

$$H_0:\theta_1=\theta_2,\ldots,\theta_k, H_1:\theta_1,\theta_2,\ldots,\theta_k\text{不全都相等}$$

在原假设成立，同一个区组内的k个观察值独立同分布。为此我们计算每一个数据在它所在区组的k个观察值中的秩。记第i个处理第j个区组的观察值x_{ij}在第j个区组的k个观察值$x_{1j},x_{2j},\ldots,x_{kj}$中的秩为$R_{ij}$。为简化讨论，不妨假设同一个区组内的$k$个观察值互不相等，从而有$R_{ij}=1,2,..,k$。同Kruskal-Wallis检验，Friedman检验的基本思想也是使用x_{ij}的秩R_{ij}代替x_{ij}，然后计算组间平方和SSB。

我们一共有k组。第i组就是第i个处理的b个观察值的秩为$\{R_{i1}, R_{i2}, \ldots, R_{ib}, i = 1, 2, \ldots, k$，这$k$组之间的组间平方和为$SSB = b\sum_{i=1}^{k}(\overline{R_i} - \overline{R})^2$，其中$\overline{R_i} = \sum_{j=1}^{b} R_{ij}/b$是第$i$个处理的$b$个观察值的秩的平均，$i = 1, 2, \ldots, k, \overline{R} = \sum_{i=1}^{k}\sum_{j=1}^{b} R_{ij}/N = (k+1)/2$是总的平均，为一个常数。

这个SSB可以作为检验统计量。我们在它的值比较的时候认为这k个位置参数不全都相等。Friedman将SSB乘以常数因子$12/(k(k+1))$，从而将SSB变换为$Q = \frac{12}{k(k+1)}SSB = \frac{12b}{k(k+1)}\sum_{i=1}^{k}(\overline{R_i} - (k+1)/2)^2$。显然，这个$Q$也可以作为检验统计量。我们在$Q$的值比较大的时候认为这$k$个位置参数不全都相等。$Q$作为检验统计量的检验称为Friedman检验。

在原假设H_0为真，即同一个区组内的k个观察值独立同分布时，若k固定，而$b \to \lim$，则Friedman检验统计量Q渐进服从$\chi^2(k-1)$分布。

随机区组试验设计资料，也可直接计算F值作F检验，F值计算步骤如下：

将每一区组的数据按大小排列，有相同数据以平均等级计算，其秩次为R_{ij}，再计算各个处理的秩和R_i，并计算所有秩的平方和：$A = \sum_{i=1}^{k}\sum_{j=1}^{b} R_{ij}^2$，及和各个处理秩和平方和的均值：$B = \frac{1}{b}\sum_{i=1}^{k} R_i^2$。其统计量$F$为：$F = \frac{(b-1)(B-bk(k+1)^2/4)}{A-B}$，其自由度$\nu 1 = k-1, \nu 2 = (b-1)(k-1)$。和Kruskal-Wallis检验一样，Friedman检验只能提示人们若干总体的中心可能不全相等，而不能指出哪些总体有着相同的中心，哪些总体存在着位置方面的差异，于是我们必须进行多重比较。

例8.11三家汽车公司油耗差异性验证 美国通用、福特与克莱斯勒汽车公司5种不同车型的某年产品油耗情况如表8-2所列：

公司 超小型 小型 中型 大型 运动型

通用 20.3 21.2 18.2 18.6 18.5

福特 25.6 24.7 19.3 19.3 20.7

克莱斯勒 24.0 23.1 20.6 19.8 21.4

数据分析关心的问题之一是三个公司汽车耗油有无差异，如果这些数据满足方差分析中 所需要的条件，我们可以直接进行方差分析进行统计检验。若在这些条件根本无法验证与确 保的情况下，则应使用非参数的Friedman检验方法。

In [ ]:

```
X=matrix(c(20.3,21.2,18.2,18.6,18.5,25.6,24.7,19.3,19.3,20.7,24.0,23.1,20.6,19.8,21.4) ,5)
friedman.test(X)
```

于是由Friedman秩和检验可知，三种汽车公司5种不同车型产品油耗不完全相同。

K-S检验

K-S单样本总体分布验证

这是一种基于经验分布函数（ECDF）的检验，记：$D = max\left|F_n(x) - F_0(x)\right|$，$F_n(x)$表示一组随机样本的累积概率函数，$F_0(x)$表示真实的分布函数。

当原假设为真时，D的值应较小，若过大，则怀疑原假设，从而对于给定的α，拒绝域为 $R = \{D > d\}, p = P\{D > d\} = \alpha$，又$p = P\{D_n \geq \hat{D}_n\}$。

设母体ξ有连续分布函数$F(x)$，从中抽取容量为n的样本，并设经验分布函数为$F_n(x)$，则 $D_n = sup_x\left|F_n(x) - F(x)\right|$在$n \to \lim$ 时有极限分布

$$P(\sqrt{n}D_n) < \lambda \to K(\lambda) = \begin{cases} \sum_{j=-\lim}^{n}(-1)^j exp(-2j^2\lambda^2) & \lambda > 0 \\ 0 & \lambda \leq 0 \end{cases}$$

例：检验是否服从已知参数的连续型分布

In [ ]:

```
x=rnorm(50)
y=runif(50,0,1)
ks.test(x,"pnorm",mean=0,sd=1)
ks.test(y,"punif",0,1)
ks.test(x,"pexp",0.5)
```

K-S两独立样本同质检验

对两组数据做(作)箱线图进行比较：

In [ ]:

```
x1=c(48,47,44,45,46,47,43,47,42,48)
x2=c(36,45,47,38,39,42,36,42,46,35)
boxplot(x1,x2,horizontal=T,names=c("x1","x2"))
```

继续用R做k-s检验：

In [ ]:

```
ks.test(x1,x2)
```

Warning message:在有连结的情况下无法正确计算p-值in:ks.test(x1,x2)由于样本量太小，不用理会渐近检验的结果。因此，对于0.05的显著性水平，不能拒绝两个分布相同的零假设。

常用正态性检验

偏度、峰度检验法

(1)用样本偏度、峰度的极限分布做检验偏度(Skewness)反映单峰分布的对称性。峰度(Kurtosis)反映分布峰的尖峭程度。为了能直观地了解两者差异，可以参看表8-3。

(2)Jarque-Bera检验(偏度和峰度的联合分布检验法)

样本偏度b_S和样本峰度b_k还可以联合起来作为正态性检验问题的检验统计量。检验统计量为 $JB = \frac{n-k}{6}(S^2 + \frac{1}{4}K^2)\ \chi^2(2)$，$JB$过大或过小时，拒绝原假设。相对来说，$JB$检验法较为实用。

J-B检验可以用使用tseries包中的jarque.bera.test函数。

例：J-B检验每加仑汽车行程数是否服从正态分布。

In [ ]:

```
#install.packages("tseries")
library(tseries)
jarque.bera.test(mtcars$mpg)
```

说明没有足够证据拒绝服从正态分布的原假设。

Shapiro-Wilk(W检验)

(1)一元正态性检验

夏皮洛—威尔克(Shapiro-Wilk)检验也称为W检验，这个检验当8≤n≤50时可以使用。过小样本n<8对偏离正态分布的检验不太有效。大样本可以使用后面介绍的K-S检验。

W检验时建在来次序统计量的基础上，将n个独立观测值按非降次序排列，记为$x_1, x_2, \ldots, x_n$，检验统计量为：

$$W = \frac{[\sum_{i=1}^{n}(a_i - \overline{a})(x_i - \overline{x})]^2}{\sum_{i=1}^{n}(a_i - \overline{a})^2(\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

其中系数$a_1, a_2, \ldots, a_n$在样本容量为n时有特定的值。

另外，系数$a_1, a_2, \ldots, a_n$还具有如下性质：
$a_i = -a_{n+1-i}, i = 1, 2, \ldots, \sum_{i=1}^{n} a_i = 0, \sum_{i=1}^{n} a_i^2 = 1$，据此可将$W$检验简化为：

$$W = \frac{[\sum_{i=1}^{n} a_i(x_{n+1-i} - x_i)]^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

可以证明，总体分布为正态分布时，W的值应该接近1，因此，在显著性水平α下，如果统计量W的值小于其α分位数，则拒绝原假设，即拒绝域为$\{W \leq w_\alpha\}$。

可以进行关于正态分布的Shapiro-Wilk检验。对汽车耗油量数据进行正态性检验。

例：验证每加仑汽车行程数是否服从正态分布。

In [ ]:

```
attach(mtcars)
shapiro.test(mtcars$mpg)
```

说明没有足够证据拒绝服从正态分布的原假设。

(2)多元正态分布检验(Shapiro-Wilk Multivariate Normality Test)

例：验证欧洲主要股指是否服从多元正态分布。在R的mvnormtest包中，数据集“EuStockMarkets”中包含了1991，1998年欧洲主要 股票市场股指收盘价(德国DAX，瑞士SMI，法国CAC，和英国FTSE).周末和假日已经略去 。这四维变量是否服从多元正态分布。(数据由ErsteBankAG，Vienna，Austria.提供。)

In [ ]:

```
#install.packages("mvnormtest")
library(mvnormtest)
data(EuStockMarkets)
dim(EuStockMarkets)
head(EuStockMarkets)
```

在R中可以调用mvnormtest包中的mshapiro.test函数进行多元正态性检验。

In [ ]:

```
C<-t(EuStockMarkets[15:29,1:4])
mshapiro.test(C)
C<-t(EuStockMarkets[14:29,1:4])
mshapiro.test(C)
R<-t(diff(t(log(C))))
mshapiro.test(R)
dR<-t(diff(t(R)))
mshapiro.test(dR)
```

其他常用正态检验

除了以上介绍的几种正态检验方法，还有其他一些方法，R中nortest包是专门做正态性 检验的包，还提供了其他检验方法。比如：

(1)AD正态性检验(ad.test)

(2)Cramer-vonMises正态性检验(cvm.test)

(3)Lilliefors正态性检验(lillie.test)

(4)Pearson卡方正态性检验(pearson.test)

(5)Shapiro-Francia正态性检验(sf.test)

以AD正态性检验(ad.test)为例，R程序如下：

In [ ]:

```
#install.packages("nortest")
library(nortest)
ad.test(rnorm(100,mean=5,sd=3))
```