第五章 统计推断之假设检验

- 何为假设检验
- 参数假设检验
- 非参数假设检验

假设检验—女士品茶

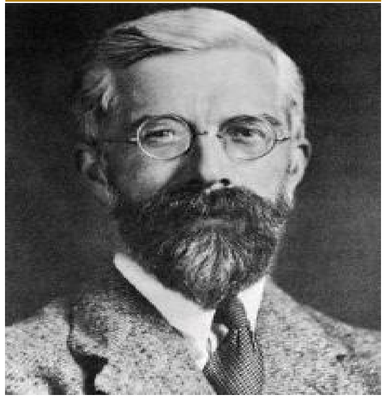
故事起因：20世纪20年代后期，在英国剑桥一个夏日的午后，一群大学的绅士和他们的夫人们，正围坐在户外的桌旁，享用着下午茶。在品茶过程中，一位女士坚称：把茶加进奶里和把奶加进茶里，做法不同，会是茶的味道品起来不同。在场的人，对这位女士的“胡言乱语”嗤之以鼻。然而，在座的一个身材矮小，戴着厚眼镜，下巴上蓄着的短尖髯开始变灰的先生，却不这么看，他对这个问题很感兴趣。他兴奋地说道：“让我们来检验这个命题吧！”

假设检验—女士品茶

故事发展：一群人就热心帮忙准备实验，准备了若干杯奶茶，有些是先放茶再加牛奶，有些先放牛奶再加茶，并将这些奶茶随机排序让这位女士品茗。在设计实验时，为了避免许多不相关的因素影响这位女士的口味辨别，还需要将茶和牛奶充分混合的时间，泡茶的时间及水的温度控制一样等等。

故事结局：在决战来临的气氛中，女士拿起第一杯茶，品了一小会儿，然后断言这一杯是先倒的茶后加的奶。然后，又拿起第二杯……，最后，竟然正确地分辨出了每一杯茶！

Ronald A. Fisher, PhD, Statistician & geneticist



Muriel Bristol-Roach, PhD, alga biologist

假设检验—基本思想

调制十杯其他条件一摸一样而仅仅是倒茶倒奶顺序相反的茶，让该女士品尝

- 假设该女士不具备区分能力 ⟸ 原假设
- 若她准确地鉴别了这十杯茶，在假设成立条件下，极其反常 ⟸ 小概率事件

小概率事件不会发生

原假设不成立

那么如果该女士鉴别出了其中的8杯，或者6杯，或者4杯，
如何判断该女士是否具备区分能力？

假设检验

假设检验—基本概念

(Hypothesis test)

概念：

- 事先对总体参数或者分布做出某种假设
- 利用样本信息来判断假设是否成立
- 主要包括参数检验和非参数检验

作用：

- 如果通过了检验，不能拒绝原假设
- 如果通不过检验，则拒绝原假设
- 只能证伪

反证法

假设检验—基本步骤

- 根据具体问题，建立原假设和备择假设
- 构造合适的统计量（使其尽可能服从已知的抽样分布），计算分布或者概率
- 给定显著水平和确定临界值
- 作出决策：如果统计量落在拒绝域，拒绝原假设

H_0：该女士不具备区分能力
H_1：该女士具备区分能力

鉴别正确的杯数/概率

临界值为2杯/显著水平为0.05

原假设（Null hypothesis)：一般研究者想收集证据予以反对的假设，表示为H_0。

备择假设（Alternative hypothesis)：一般研究者想收集证据予以支持的假设，表示为H_1。

由于假设检验中只有在小概率事件发生的情况下才拒绝原假设，因此在假设检验过程中是保护原假设的。

参数假设检验—假设

有三种形式：

总体　临界值

（1）双侧检验H_0: $\mu = \mu_0$,　H_1: $\mu \neq \mu_0$ (不等，有差异)；

（2）左侧检验H_0: $\mu \geq \mu_0$,　H_1: $\mu < \mu_0$ (降低，减少)；

若所要检验的是样本所取自的总体的参数值是否小于某个特定值时

（3）右侧检验H_0: $\mu \leq \mu_0$,　H_1: $\mu > \mu_0$ (提高，增加）

如果所要检验的是样本所取自的总体的参数值是否大于某个特定值时。

采用哪种形式要根据实际问题。

参数假设检验—假设

例1： 某种饮料的易拉罐瓶的标准容量为335毫升，为对生产过程进行控制，质量监测人员定期对某个分厂进行检查，确定这个分厂生产的易拉罐是否符合标准要求。如果易拉罐的平均容量大于或小于335毫升，则表明生产过程不正常。试陈述用来检验生产过程是否正常的原假设和备择假设。

$H_0: \mu = 335ml$

$H_1: \mu \neq 335ml$

例2： 消费者协会接到消费者投诉，指控品牌纸包装饮料存在容量不足，有欺骗消费者之嫌，包装上标明的容量为250毫升。消费者协会从市场上随机抽取50盒该品牌纸包装饮品进行假设检验，试陈述此假设检验中的原假设和备择假设。

$H_0: \mu \geq 250ml$

$H_1: \mu < 250ml$

假设检验—假设

例3： 一家研究机构估计，某城市中家庭购买有价证券的比率超过30%。为验证这一估计是否正确，该研究机构随机抽取了一个50户组成的样本进行，试陈述此假设检验中的原假设和备择假设。

$$H_0: \mu \leq 30\%$$

$$H_1: \mu > 30\%$$

假设检验—构造统计量

根据样本观测结果计算得到的，并据以对原假设和备择假设作出决策的某个样本统计量

总体分布	样本容量	σ已知	σ未知
正态分布	大样本	$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim N(0,1)$
	小样本*	$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)$
非正态分布	大样本	$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim N(0,1)$

- 常用的抽样分布构造的统计量

H_0为真时

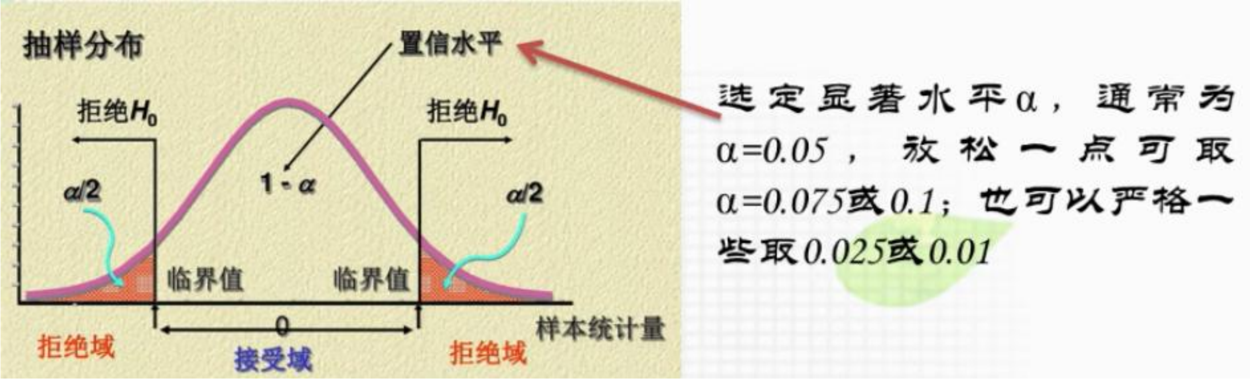
(1) $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

(2) $\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)$

(3) $\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \sim \chi^2(n)$

(4) $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 \sim \chi^2(n-1)$

假设检验——显著水平



统计量 $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$P\left\{\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \le Z_{\alpha/2}\right\} = 1 - \alpha$$

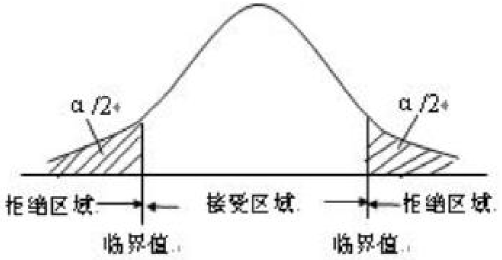
$$P(|X| \ge u) \quad vs \quad \alpha$$

- 显著性水平用α表示；
- 在假设检验中,显著性水平的含义是当原假设正确时却被拒绝的概率,即假设检验中犯弃真错误的概率。它是由人们根据检验的要求确定的。

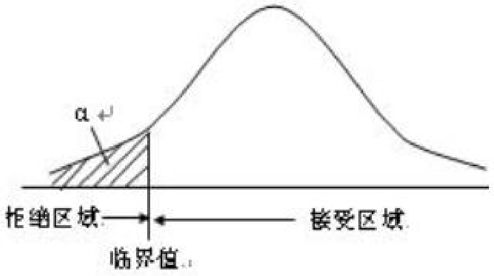
如果统计量的值正好落在拒绝域之内（包括边界），那么拒绝原假设；反之不能拒绝原假设。

假设检验—拒绝域

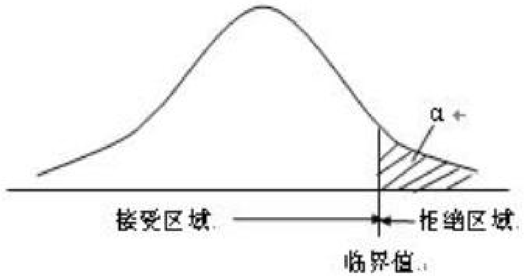
由显著水平和分布决定



（a） 双侧检验



（b） 左侧检验



（c） 右侧检验

（1）双侧检验H_0: $\mu = \mu_0$， H_1: $\mu \neq \mu_0$ (不等，有差异)；

（2）左侧检验H_0: $\mu \geq \mu_0$， H_1: $\mu < \mu_0$ (降低，减少)；

若所要检验的是样本所取自的总体的参数值是否小于某个特定值时

（3）右侧检验H_0: $\mu \leq \mu_0$， H_1: $\mu > \mu_0$ (提高，增加）

如果所要检验的是样本所取自的总体的参数值是否大于某个特定值时。

如果统计量的值正好落在拒绝域之内（包括边界），那么拒绝原假设；反之不能拒绝原假设。

假设检验—实例1：双边检验

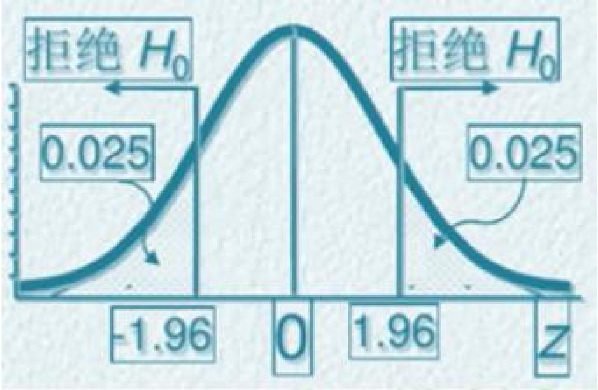
例1： 一种罐装饮料采用自动生产线生产，每罐的容量是255ml，标准差为5ml，服从正态分布，为检验每罐容量是否符合要求，质检人员在某天生产的饮料中随机抽取了16罐进行检验，测得每罐平均容量为257.2ml。取显著水平α =0.05，检验该天生产的饮料容量是否符合标准要求？

假设：$H_0: \mu = 255$；$H_1: \mu \neq 255$

已知：$\alpha = 0.05$； $n = 16$

统计量：$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{257.2-255}{5/\sqrt{16}} = 1.76$

确定拒绝域：



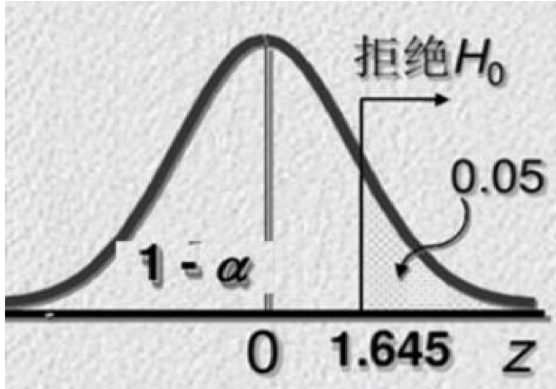
决策：不能拒绝H_0

结论分析：当天生产饮料的容量与标准容量没有显著差异，其差异是由随机因素引起的……

假设检验—实例2：单边检验

例2： 一种罐装饮料采用自动生产线生产，每罐的容量是255ml，标准差为5ml，服从正态分布。换了一批工人后，质检人员在某天生产的饮料中随机抽取了16罐进行检验，测得每罐平均容量为257.2ml。取显著水平α =0.05，检验该天生产的饮料容量是否增加了？

右侧检验



假设： $\boldsymbol{H_0: \mu \leq 255; H_1: \mu > 255}$

统计量： $z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{257.2-255}{5/\sqrt{16}} = 1.76$

> 例1：
> $H_0: \mu = 255$； $H_1: \mu \neq 255$
> 决策：不能拒绝H_0

确定拒绝域： $\boldsymbol{\alpha = 0.05}$； $\boldsymbol{n = 16}$

决策： 拒绝H_0

结论分析： 当天生产饮料的容量与标准容量有显著差异，可以认为换工人后容量增加了……

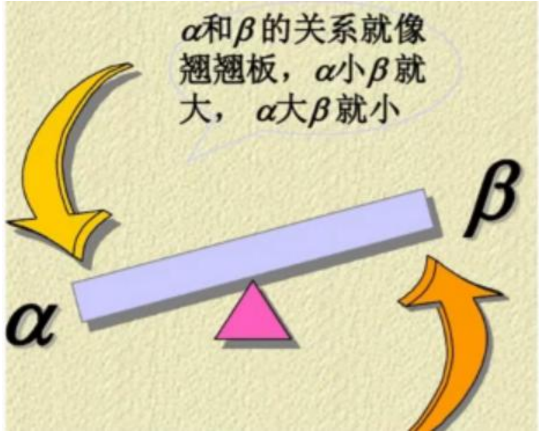
假设检验—两类错误

	实际情况	
决定	H_0为真	H_0不真
拒绝H_0	第一类错误 (Type I error) 弃真	正确
接受H_0	正确	第二类错误 (Type II error) 取伪

犯两类错误的概率：

- $P\{拒绝H_0|H_0为真\} = \alpha$
- $P\{接受H_0|H_0不真\} = \beta$

- 显著性水平α为犯第一类错误的概率；
- α 和 β 的关系就像跷跷板，α 小 β 就大，α 大 β 就小。



如何减免：

1. 你不能同时减少两类错误！
2. 为了保护原假设，尽可能避免第一类错误；
3. 增加样本容量，减少犯第二类错误的概率；

5.2 参数假设检验

- 单个总体
- 两个总体

假设检验—单个总体：均值检验

1. 方差已知情形：

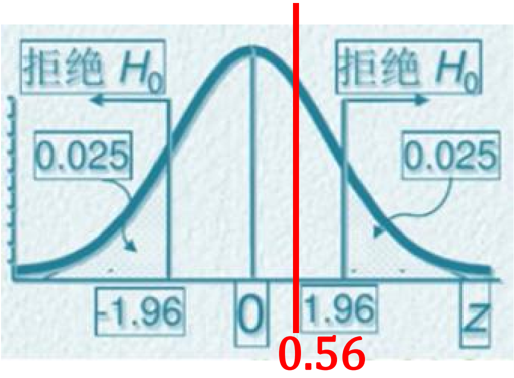
想拒绝的是：轮轴均值就是5

某公司生产轮轴，直径均值为5.00cm，假定轮轴的直径服从正态分布，标准差为1.00cm。目前你作为该公司的检验员，通过抽取的样本$(X_1,X_2,X_3,\ldots,X_n)$，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

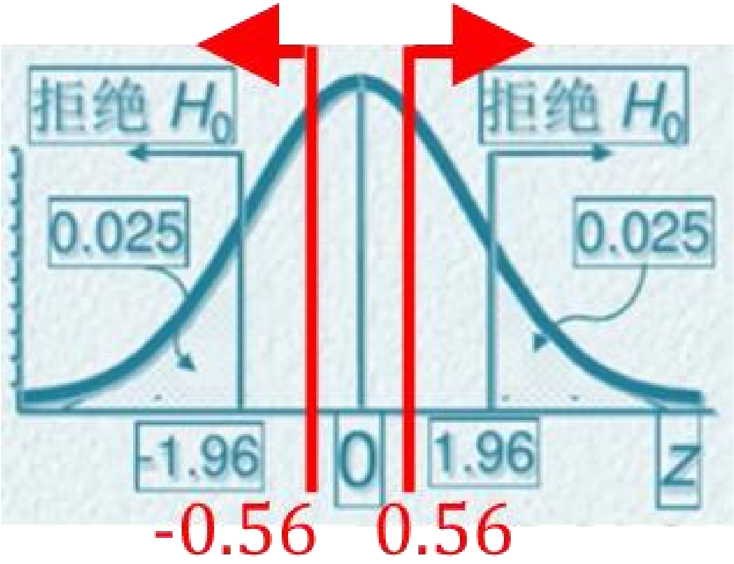
- 原假设H_0：$\mu=\mu_0=5$
- 备择假设H_1：$\mu\neq\mu_0$
- 构造统计量：$Z=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$

$$Z=\frac{5.126-5}{1/\sqrt{20}}=0.56$$



结论：不能拒绝原假设，轮轴直径的均值是5cm

- 构造统计量： $Z=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$ $\qquad Z=\dfrac{5.126-5}{1/\sqrt{20}}=0.56$



$P(|X|\geq u)\quad vs\quad \alpha$

从概率的角度，还可以计算统计量取值的概率（拒绝原假设犯错误的概率）*P*。

- 如果 *P* 小于或等于显著水平，表明小概率事件发生了，因此拒绝原假设；
- 反之不能拒绝原假设。

p-value:原假设为真时，比样本更极端情况出现的概率，即拒绝原假设出错的概率。

P值越小，表明小概率事件发生，则越倾向拒绝原假设。

假设检验—单个总体：均值检验

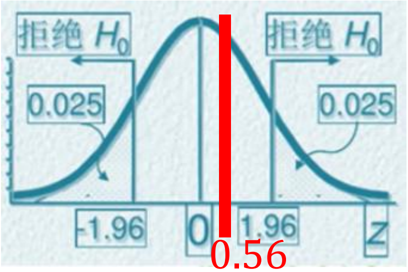
想拒绝的是：轮轴均值就是5

1. 方差已知情形：

某公司生产轮轴，直径均值为5.00cm，假定轮轴的直径服从正态分布，标准差为1.00cm。目前你作为该公司的检验员，通过抽取的样本$(X_1, X_2, X_3, \dots, X_n)$，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

- 原假设H_0：$\mu = \mu_0 = 5$
- 备择假设H_1：$\mu \neq \mu_0$
- 构造统计量：$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$



$$Z = \frac{5.126-5}{1/\sqrt{20}} = 0.56$$

$P(|X| \geq u) \quad vs \quad \alpha$

```
> z.test<-function(x,mu0,sigma){
+   z=(mean(x)-mu0)/(sigma/sqrt(length(x)))
+   p=2*(1-pnorm(abs(z)))
+   return(list(z=z,p=p))
+ }
> x=c(4.89,5.99,5.89,6.22,4.79,5.47,4.5,6.61
4,5.27,4.48,4.05)
> z.test(x,5,1)
$z
[1] 0.5657252

$p
[1] 0.5715806
```

```
p=2*pnorm(abs(z),lower.tail = FALSE)
```

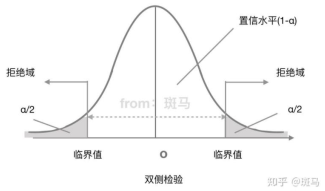
结论：不能拒绝原假设，轮轴直径的均值是5cm

假设检验—单个总体：均值检验

```
> z.test<-function(x,mu0,sigma,alternative="twoside"){
+   z=(mean(x)-mu0)/(sigma/sqrt(length(x)))
+   if(alternative=="twoside"){
+     #p=2*(1-pnorm(abs(z)))
+    p=2*pnorm(abs(z),lower.tail = FALSE)
+
+   }
+  else if (alternative=="less"){
+    p=pnorm(z)
+  }
+   else if(alternative=="greater"){
+    p=1-pnorm(z)
+   }
+   return(list(z=z,p=p))
+ }
```

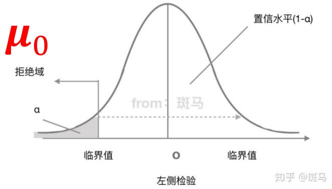
H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$

双侧检验



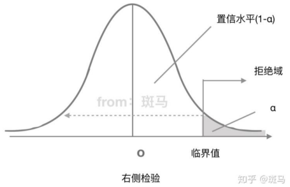
H_0: $\mu \geq \mu_0$; H_1: $\mu < \mu_0$

左侧检验



右侧检验

H_0: $\mu < \mu_0$; H_1: $\mu \geq \mu_0$



假设检验—单个总体：均值检验

1. 方差已知情形：

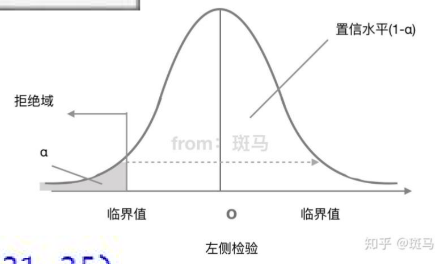
某汽车生产商声称其生产的汽车每加仑汽油可行驶的里程不低于25英里，标准差为2.4英里。消协组织了一个由10位汽车主组成的小组，他们的汽车每加仑汽油的可行驶英里数如下表。假定汽车每加仑可行驶里程服从正态分布。则汽车生产商的诺言可信吗？

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

```
> z.test<-function(x,mu0,sigma,alternative="twoside"){
+   z=(mean(x)-mu0)/(sigma/sqrt(length(x)))
+   if(alternative=="twoside"){
+      #p=2*(1-pnorm(abs(z)))
+     p=2*pnorm(abs(z),lower.tail = FALSE)
+
+   }
+  else if (alternative=="less"){
+    p=pnorm(z)
+  }
+   else if(alternative=="greater"){
+     p=1-pnorm(z)
+   }
+   return(list(z=z,p=p))
+ }
```

左侧检验

H_0: $\mu \geq 25$；H_1: $\mu < 25$




```
> x=c(22,24,21,24,23,24,23,22,21,25)
> z.test(x,25,2.4,alternative="less")
$z
[1] -2.766993

$p
[1] 0.002828799
```

结论：拒绝原假设，汽车生产商的诺言不可信

假设检验—单个总体：均值检验

1. 方差已知情形：

想拒绝的是：生产的汽车每加仑汽油可行驶的里程低于25英里

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

```
> z.test<-function(x,mu0,sigma,alternative="twoside"){
+   z=(mean(x)-mu0)/(sigma/sqrt(length(x)))
+   if(alternative=="twoside"){
+      #p=2*(1-pnorm(abs(z)))
+     p=2*pnorm(abs(z),lower.tail = FALSE)
+
+   }
+  else if (alternative=="less"){
+     p=pnorm(z)
+  }
+   else if(alternative=="greater"){
+     p=1-pnorm(z)
+   }
+   return(list(z=z,p=p))
+ }
```

右侧检验

H_0: $\mu \le 25$； H_1: $\mu > 25$

```
> x = c(22, 24, 21, 24, 23, 24, 23, 22, 21, 25)
> z.test(x, 25, 2.4, alternative = "greater")
$z
[1] -2.766993

$p
[1] 0.9971712
```

结论：不能拒绝原假设，
汽车生产商的诺言不可信

假设检验—单个总体：均值检验

1. 方差已知情形

设$x_1, x_2, \ldots, x_n$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）统计量 $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

（4）计算值对应的p值

（5）若 Z 值落在拒绝域内，则拒绝H_0，反之不能拒绝H_0

假设检验—单个总体：均值检验

2. 方差未知情形

设$x_1, x_2, \dots, x_n$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）统计量 $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$ 服从自由度为 $n-1$的t 分布t(n−1)

（4）计算值对应的p值

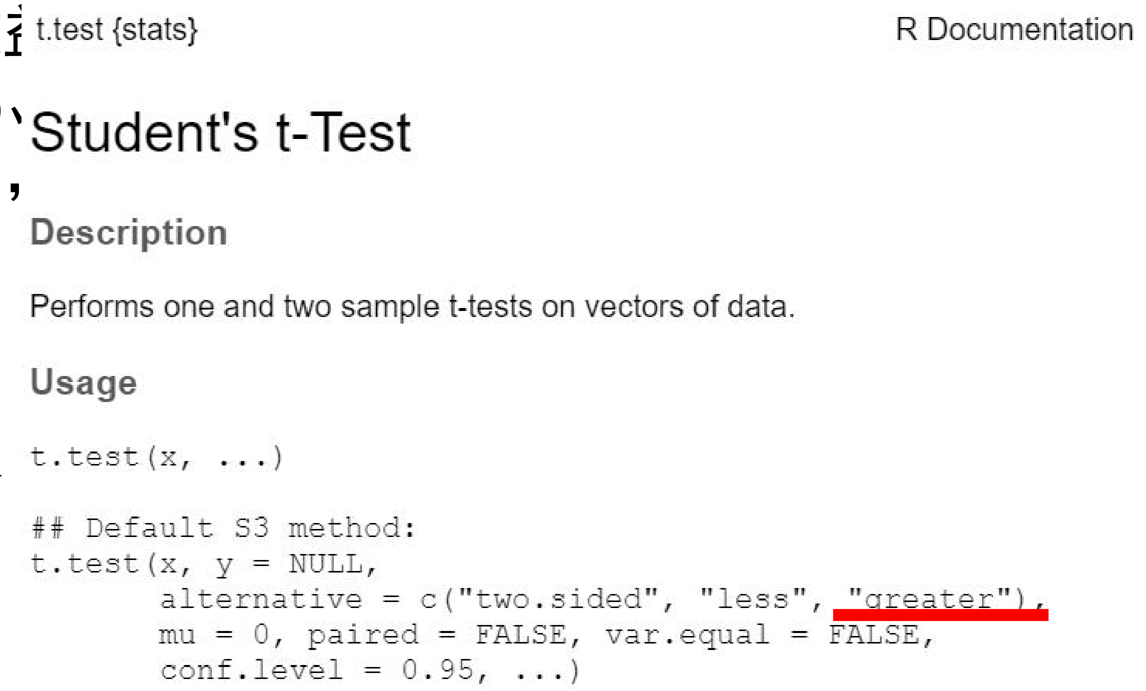
（5）若 t 值落在拒绝域内，则拒绝H_0，反之不能拒绝H_0

假设检验—单个总体：均值检验

2. 方差未知情形

想拒绝的是：新的资产管理公司平均收益率比原来资产管理公司的小

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 0.6%、49.8%、51.2%、49.7%、51.5%、 客户收益率的分布比较近似于正态分布， 来的资产管理公司？


t.test {stats} R Documentation

Student's t-Test

Description

Performs one and two sample t-tests on vectors of data.

Usage

```
t.test(x, ...)

## Default S3 method:
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

- H_0: $\mu \le 50\%$，
- H_1: $\mu > 50\%$，右侧检验
- 显著性水平为95%

假设检验—单个总体：均值检验

2. 方差未知情形

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？

- H_0: $\mu \le 50\%$，H_1: $\mu > 50\%$，右侧检验
- 显著性水平为95%

```
> t.test(income, mu=0.50,alternative="greater")

        One Sample t-test

data:  income
t = 2.9564, df = 11, p-value = 0.006529
alternative hypothesis: true mean is greater than 0.5
95 percent confidence interval:
 0.5020609       Inf
sample estimates:
mean of x
  0.50525
```

结论：拒绝原假设，
新资产管理公司的平均收益大于原来的资产管理公司

练 习

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$N(\mu,\sigma^2)$,设分别在下列条件下检验假设（α =0.05）

$$H_0\colon \mu = 100 \ \ vs\ H_1\colon \mu > 100$$

（I） 已知σ =0.05

（2） σ 未知

5.2 参数假设检验

- 单个总体
- 两个总体

假设检验—双样本

1. 工科211本科在读。男多女少环境内，年级前十有八个女生；
2. 大四那年，年级有13个女生，保研9个，男生只有2个保研；
3. 加拿大一对夫妇研究了来自308家全球机构的369个样本（包含53.8万男生和59.5万女生）中小学到研究生的成绩。结论：女生成绩好于男性已持续了近百年，差距在理工科和高学历时缩小。

社会因素和文化因素是主要原因，且游戏，玩具和社交平台，都更吸引男生而非女生的注意力。

“阴盛阳衰”成趋势，男生逐渐成为“弱势群体”，你同意吗？

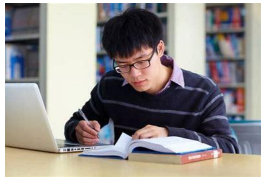
假设检验—双样本

假设$X_1, X_2, \ldots, X_{n_1}$是来自总体$X \sim N(\mu_1, \sigma_1^2)$的样本， $Y_1, Y_2, \ldots, Y_{n_2}$是来自总体$Y \sim N(\mu_2, \sigma_2^2)$的样本，且两样本独立。其检验问题有：

双边检验：　$H_0: \mu_1 = \mu_2$　$H_1: \mu_1 \neq \mu_2$

右边检验 1：　$H_0: \mu_1 \leq \mu_2$　$H_1: \mu_1 > \mu_2$

左边检验 2：　$H_0: \mu_1 \geq \mu_2$　$H_1: \mu_1 < \mu_2$



(1) 方差 σ_1^2 和 σ_2^2 已知，由统计知识 (1.5.4 节的式 (1.97)) 可知，当 H_0 为真时，

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

因此，当 Z 满足 (称为拒绝域)

双边检验：　$|Z| \geq Z_{\alpha/2}$,

单边检验 I:　$Z \geq Z_\alpha$,

单边检验 II:　$Z \leq -Z_\alpha$.

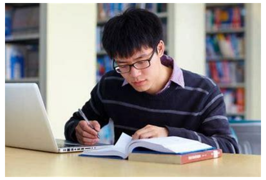
假设检验—双样本

假设$X_1, X_2, \ldots, X_{n_1}$是来自总体$X \sim N(\mu_1, \sigma_1^2)$的样本， $Y_1, Y_2, \ldots, Y_{n_2}$是来自总体$Y \sim N(\mu_2, \sigma_2^2)$的样本，且两样本独立。其检验问题有：

双边检验： $H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2$

右边检验 1： $H_0: \mu_1 \leq \mu_2 \quad H_1: \mu_1 > \mu_2$

左边检验 2： $H_0: \mu_1 \geq \mu_2 \quad H_1: \mu_1 < \mu_2$



(2) 方差 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知. S_1^2 和 S_2^2 分别是 X 和 Y 的样本方差. 由统计知识 (1.5.4 节的式 (1.98)) 可知, 当 H_0 为真时,

$$T = \frac{\overline{X} - \overline{Y}}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

其中

$$S_w = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}.$$

因此, 当 T 满足 (称为拒绝域)

双边检验: $|T| \geq t_{\alpha/2}(n_1 + n_2 - 2)$,

单边检验 I: $T \geq t_\alpha(n_1 + n_2 - 2)$,

单边检验 II: $T \leq -t_\alpha(n_1 + n_2 - 2)$.

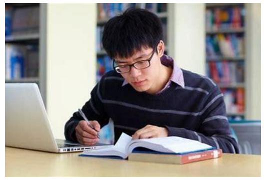
假设检验—双样本

假设$X_1, X_2, \ldots, X_{n_1}$是来自总体$X \sim N(\mu_1, \sigma_1^2)$的样本， $Y_1, Y_2, \ldots, Y_{n_2}$是来自总体Y$\sim N(\mu_2, \sigma_2^2)$的样本，且两样本独立。其检验问题有:

双边检验: $\mathrm{H}_0: \mu_1 = \mu_2 \quad \mathrm{H}_1: \mu_1 \neq \mu_2$

右边检验 1: $\mathrm{H}_0: \mu_1 \leq \mu_2 \quad \mathrm{H}_1: \mu_1 > \mu_2$

左边检验 2: $\mathrm{H}_0: \mu_1 \geq \mu_2 \quad \mathrm{H}_1: \mu_1 < \mu_2$



(3) 方差 $\sigma_1^2 \neq \sigma_2^2$ 未知. S_1^2 和 S_2^2 分别是 X 和 Y 的样本方差.

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(\widehat{\nu}) \qquad \widehat{\nu} = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2 \Big/ \left(\frac{(S_1^2)^2}{n_1^2(n_1 - 1)} + \frac{(S_2^2)^2}{n_2^2(n_2 - 1)}\right)$$

假设检验—两独立样本

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，这些数据能够提供充分的证据说明开车和坐公交上班时间所用时间相同吗？用显著水平5%，并考虑单尾检测还是双尾检测。

公交：48，47，44，45，46，47，43，47，42，48

开车：38，45，47，38，39，42，36，42，46，35

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

两方差是否已知？ 两方差是否相等？

```
> bus<-c(48,47,44,45,46,47,43,47,42,48)
> car<-c(38,45,47,38,39,42,36,42,46,35)
> var(bus)
[1] 4.455556
> var(car)
[1] 17.95556
```

```
> var.test(bus,car)

        F test to compare two variances

data:  bus and car
F = 0.24814, num df = 9, denom df = 9, p-value = 0.04985
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.06163535 0.99902454
sample estimates:
ratio of variances
         0.2481436
```

方差不相等

假设检验—两独立样本

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，这些数据能够提供充分的证据说明开车和坐公交上班时间所用时间相同吗？用显著水平5%，并考虑单尾检测还是双尾检测。

公交：48，47，44，45，46，47，43，47，42，48

开车：38，45，47，38，39，42，36，42，46，35

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

两方差是否已知？ 两方差是否相等？

```
> bus<-c(48,47,44,45,46,47,43,47,42,48)
> car<-c(38,45,47,38,39,42,36,42,46,35)
> var(bus)
[1] 4.455556
> var(car)
[1] 17.95556
```

```
> t.test(bus,car,var.equal = F)

        Welch Two Sample t-test

data:  bus and car
t = 3.2731, df = 13.208, p-value = 0.00594
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.671012 8.128988
sample estimates:
mean of x mean of y 
     45.7      40.8 
```

时间不相等

结论：拒绝原假设，
公交车与汽车所用时间不等

假设检验—两独立样本

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，这些数据能够提供充分的证据说明开车和坐公交上班时间所用时间谁更少吗？用显著水平5%，并考虑单尾检测还是双尾检测。

公交：48，47，44，45，46，47，43，47，42，48

开车：38，45，47，38，39，42，36，42，46，35

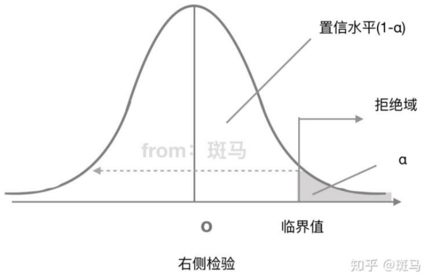
H_0: $\mu_{bus} \le \mu_{car}$　H_1: $\mu_{\text{bus}} > \mu_{car}$

```
> bus<-c(48,47,44,45,46,47,43,47,42,48)
> car<-c(38,45,47,38,39,42,36,42,46,35)
```

```
> t.test(bus,car,var.equal = F,alternative="greater")

        Welch Two Sample t-test

data:  bus and car
t = 3.2731, df = 13.208, p-value = 0.00297
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.252034      Inf
sample estimates:
mean of x mean of y 
     45.7      40.8 
```



结论：拒绝原假设，
汽车所用时间更短

假设检验—两配对样本

- ✓ 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。

假设检验—两配对样本

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

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

在正态性假定下，$d = X - Y$ 近似服从 $N\left(\mu, \sigma_{\mathrm{d}}^2\right)$。其中，$\mu = \mu_1 - \mu_2$，$\sigma_{\mathrm{d}}^2 = \sigma_1^2 + \sigma_2^2$需要比较 μ_1 与 μ_2 大小的问题转变成了 μ是否为0

假设检验—两配对样本

对差值 d 是否为0进行检验：

（1）检验假设: $H_0: d = 0$ 和 $H_A: d \neq 0$

（2）给出显著性水平α

（3）统计量： $t = \overline{d}/$（$s_d/\sqrt{n}$），其中$d_i = x_i - y_i$，$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$， $s_d = [\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2]^{\frac{1}{2}}$

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

（4）计算 t值对应的 p 值

（5）若 t 值落在拒绝域内，则拒绝 H_0，反之不能拒绝 H_0

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

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

```
> 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 = 14.164, 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
```

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

$d = \mu_1 - \mu_2$

$H_0: d \leq 8.5$, $H_A: d > 8.5$

```
> 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,alternative="greater",mu=8.5,paired = T)

        Paired t-test

data:  before and after
t = 1.9413, df = 9, p-value = 0.04207
alternative hypothesis: true difference in means is greater than 8.5
95 percent confidence interval:
 8.575214      Inf
sample estimates:
mean of the differences
                   9.85
```

假设检验—两配对样本

t.test {stats} R Documentation

Student's t-Test

Description

Performs one and two sample t-tests on vectors of data.

Usage

```
t.test(x, ...)

## Default S3 method:
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

单样本还是双样本？输入X和Y样本数据

根据定义备择假设

mu=X-Y

$\boxed{\text{paired = FALSE,}}$

两个总体方差是否相等

置信水平

是否是配对样本

- 假设总体服从xx分布已知

$$\text{e.g., } f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

参数假设检验

如果总体的分布未知呢?

想要反对的假设

H_0： 原假设
H_1： 备择假设

构造统计量
判断其分布

给出显著水平α

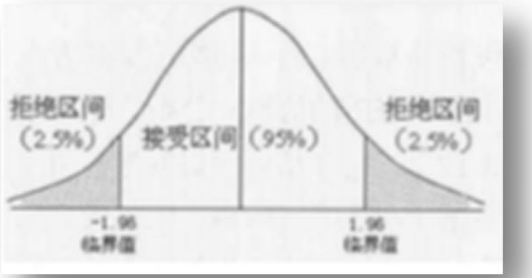
在原假设成立时，
判断统计量所在区域，
计算p值

给出结论

$$\mu=\mu_0$$

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

$$\frac{\bar{x}-\mu}{s/\sqrt{n}}\sim t(n-1)$$



$P(|X|\geq u)$ vs α

原假设成立时，比样本更极端情况发生的概率，即拒绝原假设犯错误的概率

假设检验的应用——AB测试